UNIVERSIDAD DE MURCIA

FACULTAD DE LETRAS

Identification and Analysis of the Specialised Vocabulary of British Law Reports: A Corpus-driven Study of this Legal Genre at the Core of Common Law Legal Systems

Identificación y Análisis del Vocabulario Especializado de los Repertorios de Jurisprudencia Británicos: Estudio Basado en un Corpus de este Género Legal, Fundamento de los Sistemas Legales *Common Law*

Dª María José Marín Pérez
2013

To Antonio, María and Antonio.
Each and every word of these is yours.

Between two hawks, which flies the higher pitch;

Between two dogs, which hath the deeper mouth;

Between two blades, which bears the better temper;

Between two horses, which doth bear him best;

Between two girls, which hath the merriest eye;

I have perhaps some shallow spirit of judgement.

But in these nice sharp quillets of the law,

Good faith, I am no wiser than a daw.

1 Henry VI. (2.4.14-20).

William Shakespeare (1623).

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisor, Dr. Aquilino Sánchez, for his invaluable help; for pushing me one step further whenever I thought it was not possible. I must thank him for his professionalism in the way that he has dealt with my work and the astonishing speed he has always reviewed it with.

I also wish to thank Dr. Camino Rea for her unconditional support and her friendship. Without her work, mine would have not been possible. She led me into the field of corpus linguistics and taught me the very basics of a discipline I was completely ignorant of. She has always been present throughout this hard process both on an academic and, above all, personal level and I must thank her for that.

My grateful thanks are extended to Dr. Juan Manuel Sáez, from the University of Alicante. Without his advice on technical matters, most of this research could have not been carried out either. I must thank him for the time he has generously dedicated to observe and analyse the linguistic data I have given him and translate it into mathematical language. But, most of all, I would like to say *thank you* for the kindness and patience he has shown to me. Regardless of the difficulties he might have been going through, I could always count on him.

I also wish to thank my husband, Antonio, for his love and care and his dedication to our children while I was busy trying to make sense of all of this. He has been my major support and the one who has understood and encouraged me to go on whenever I felt this was going nowhere. And of course, thanks to my children who, instead of becoming an obstacle, have been my inspiration.

Finally, I would like to express my gratefulness to my sister, Marta, and my parents, Trini and Santiago. I am who I am because of you.

INDEX

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ATR: Automatic Term Recognition

*AWL***:** *Academic Word List*

*BLaRC***:** *British Law Report Corpus*

*BNC***:** *British National Corpus*

CL: Corpus Linguistics

CT: Candidate Term

DDL: Data-driven learning

EAP: English for Academic Purposes

ESP: English for Specialised Purposes

*GSL***:** *General Service List*

*LACELL***:** *Lingüística Aplicada Computacional, Enseñanza de Lenguas y Lexicografía*

LC: Lexical constellation

MI: *Mutual Information*

MWT: Multi-Word Term

RC: Reference Corpus

SC: Study Corpus

SWT: Single-Word Term

TT: True Term

UKSCC: *United Kingdom Supreme Court Corpus*

RESUMEN EN ESPAÑOL

INTRODUCCIÓN

La comprensión de los términos en un texto especializado sin duda contribuye a una mejor interpretación del texto mismo. Así pues, tener acceso a estos términos puede convertirse en una ventaja para el investigador interesado en estudiar las características y las normas que gobiernan el léxico del lenguaje de especialidad.

Diversos especialistas destacan que el inglés jurídico (también conocido como *legalese*), es una variedad del inglés particularmente compleja y elaborada. D. Mellinkoff, uno de los primeros estudiosos del inglés jurídico afirma que éste tiende a ser oscuro, retórico, pomposo y aburrido (Mellinkoff, 1963: 63). La presencia de préstamos latinos, expresiones provenientes del francés antiguo, sinónimos, arcaísmos y redundancia, además del uso extendido de "palabras comunes con significados inusuales" (Mellinkoff, 1963*:* 11) caracterizan su léxico.

La mayor parte de los trabajos dedicados a la descripción del inglés jurídico (Mellinkoff, 1963; Alcaraz, 1994, 2001; Tiersma, 1999; Borja, 2000) son mayoritariamente prescriptivos pues están basados en la intuición del autor o bien en ejemplos reducidos de esta variedad del inglés.

Se ha discutido ampliamente sobre las ventajas y desventajas del uso de corpus para la descripción del funcionamiento de la lengua (Sinclair, 1991; Sánchez et al., 1995; McEnery and Wilson, 1996; Dudley-Evans and St. John, 1998; Kennedy, 1998; McEnery et al., 2006; Tognini-Bonelli, 2001; Meyer, 2002; Gries and Wulff, 2010; Cheng, 2011; etc.). La distinción chomskiana entre *competence* y *performance* está en la base de las críticas contra el uso de corpus con ese fin debido a que, según esta corriente de pensamiento, los ejemplos intuitivos generados desde el conocimiento

tácito de la lengua son los que realmente pueden emplearse para hacer generalizaciones sobre ella.

Sin embargo, los corpora han crecido en la última década hasta alcanzar dimensiones impensables en los años 50 y 60 tales como *enTenTen,* de 12.000 millones de palabras, accesible a través de herramientas online como *Sketch Engine* (Kilgariff et al., 2003). Asimismo, el desarrollo de herramientas para su procesamiento facilita enormemente el acceso a cantidades ingentes de información. Como consecuencia, las teorías sobre la lengua basadas en la descripción de los corpora son mucho más fiables que lo fueran hace 50 años cuando Chomsky rechazaba su uso como fuente sólida de información lingüística.

En el estudio del vocabulario especializado, la información que pueden proporcionar los corpora específicos es de gran valor. No obstante, en el área del inglés jurídico el número y la disponibilidad de éstos es muy reducido. Por este motivo se diseña y compila *BLaRC* (*British Law Report Corpus*), un corpus legal de sentencias judiciales de 8,85 millones de palabras con el fin de ser utilizado como fuente de vocabulario especializado para su posterior análisis.

OBJETIVOS, METODOLOGÍA Y ESTRUCTURA

El objetivo principal de esta tesis doctoral es la identificación y posterior análisis de los términos legales de *BLaRC,* un corpus de sentencias judiciales del Reino Unido. El capítulo 2 presenta y justifica sus características principales y considera cuestiones fundamentales en el diseño y la compilación del corpus tales como la relevancia de los repertorios de jurisprudencia (o *law reports*) en los sistemas legales *common law,* el tamaño ideal del corpus, modalidad de los textos, cronología de éstos, variedad léxica, etc.

El inglés jurídico es una variedad del inglés que abarca gran cantidad de géneros. Por un lado, como afirma Orts (2009), documentos de carácter público tales como las sentencias judiciales, leyes, decretos, etc. y por otro los pertenecientes al derecho privado como escrituras, testamentos, poderes notariales, acuerdos de divorcio, contratos, etc.

Como consecuencia, hubo que reducir el número de géneros en los que basar el corpus centrándonos en el derecho público, dada su relevancia. Dentro del derecho público hay dos fuentes de las que emana el derecho en los países *common law* como el Reino Unido, una de ellas son las leyes que se aprueban en el parlamento (que han ido ganando relevancia en los últimos 150 años) y la otra, la más importante, los casos ya resueltos con anterioridad por instancias generalmente superiores que han sentado precedente (siguiendo el principio de *stare decisis*), esto es, la jurisprudencia.

Las sentencias judiciales se recogen en repertorios de jurisprudencia o *law reports* que los profesionales del derecho anglosajón emplean como fuente fundamental de información para la argumentación, defensa o resolución de casos. Las sentencias recogen todo tipo de vocabulario legal perteneciente a las distintas ramas del derecho del mismo modo que se incluyen citas de las leyes y decretos aprobados en el parlamento cuando resulta pertinente.

Su riqueza léxica es, por consiguiente, innegable. Así pues, *BLaRC* está formado por sentencias judiciales dictadas por tribunales de todo el Reino Unido pertenecientes a todos los niveles de la jerarquía judicial: Tribunal Supremo, Tribunal de Apelación, Alto Tribunal, Magistrates' Courts, Tribunals, etc. y a todas las áreas del país con competencias judiciales e instituciones independientes: Irlanda del Norte, Inglaterra y Gales y Escocia, además de aquellos países de la Commonwealth que aún utilizan el Privy Council del Reino Unido como tribunal de último recurso (Bahamas, Jamaica,

Nueva Zelanda o Bermudas, entre otros). Las sentencias se agrupan por tribunales y por lo tanto por áreas del derecho a excepción del Tribunal Supremo y el Privy Council que tienen competencias y jurisdicción a todos los niveles.

Una vez compilado el corpus se procede a la selección de los métodos de reconocimiento automático de términos (métodos ATR) más efectivos en la identificación de términos tanto mono-léxicos como poli-léxicos en el capítulo 3. Se evalúan diez métodos diferentes midiendo en cada caso la precisión general alcanzada, esto es, el porcentaje de término reales identificados respecto del total de candidatos a términos extraídos. Los métodos evaluados son los siguientes:

1) *TermoStat* (Drouin, 2003)

2) Chung (2003): *Frequency ratio*

3) Kit y Liu (2008): *Rank difference*

4) *Term Frequency-Inverse Term Frequency* (*TF-IDF*) (Spark Jones, 1972)

5) *Residual Inverse Document Frequency* (*RIDF*) (Church and Gale, 1995)

6) *Keywords* (Scott, 2008)

7) *Terminus 2.0* (Nazar y Cabré, 2012)

8) *C-value* (Frantzi et al., 1999)

9) *Termextractor* (Sclano and Velardi, 2007)

10) *Textract* (Park et al., 2002)

Igualmente, se calcula la precisión cumulativa para los 2.000 primeros candidatos a término mono-léxicos y los primeros 1.400 poli-léxicos con el fin de determinar cómo afecta al nivel de precisión el aumento del número de candidatos, como se recoge en las figuras 1 y 2 respectivamente. Este cálculo se realiza incrementando el número de candidatos de manera progresiva de 200 en 200 y calculando el porcentaje de términos reales por tramos.
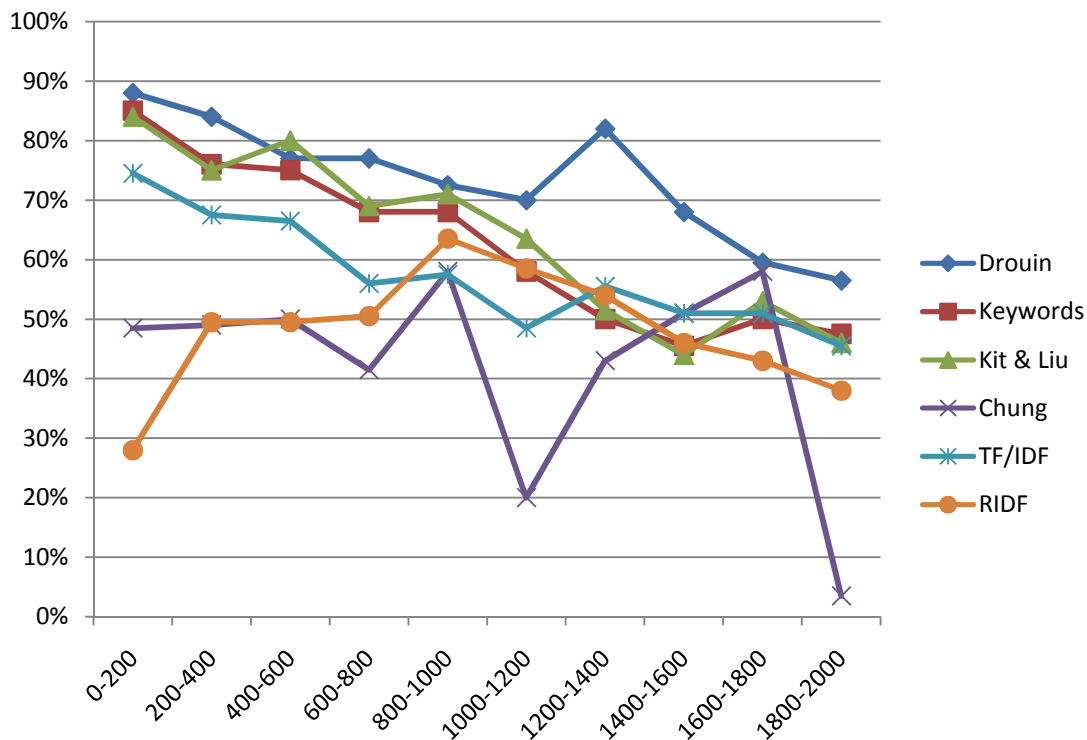
Figure 1. *Precisión cumulativa alcanzada en la identificación de términos mono-léxicos*
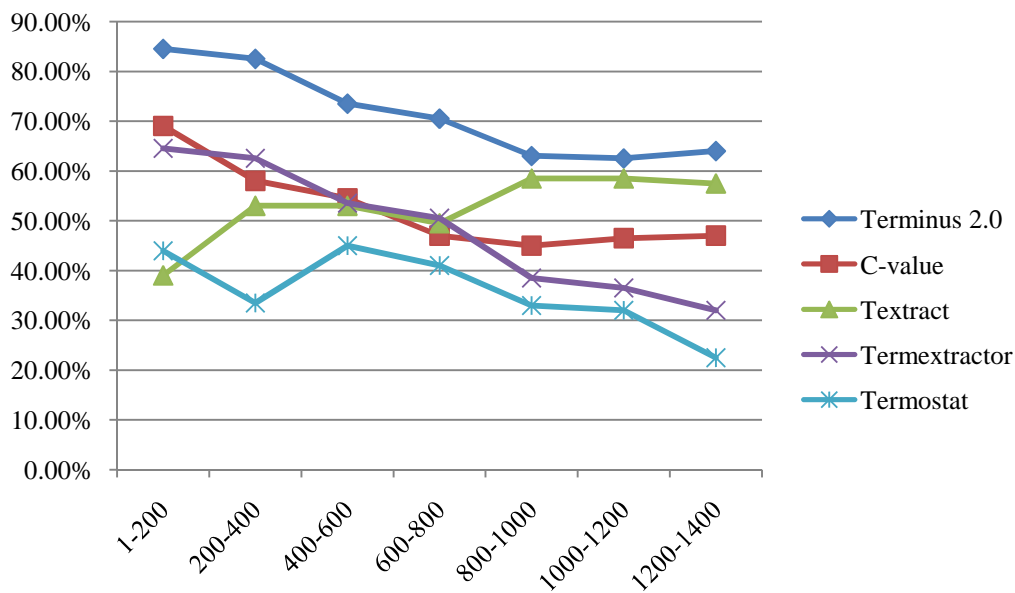
Figure 2. *Precisión cumulativa alcanzada en la identificación de términos poli-léxicos*

La evaluación de los métodos seleccionados se lleva a cabo a través de la comparación de las listas de candidatos a término con un *gold standard*, esto es, un glosario legal electrónico de 10.088 entradas, en una hoja excel para determinar qué cantidad de candidatos son términos reales en función de la coincidencia con los términos recogidos en el glosario. Para facilitar el proceso de implementación de todos los métodos ATR y su evaluación, dado el tamaño de *BlaRC,* se opta por el uso de un corpus piloto extraído de este último. Se trata de *UKSCC* (*United Kingdom Supreme Court Corpus*) un corpus de 2.6 millones de palabras que contiene sentencias judiciales emitidas por el Tribunal Supremo del Reino Unido entre los años 2008 y 2010.

El resultado de la evaluación nos lleva a la selección de *TermoStat* (Drouin, 2003) y *Terminus 2.0* (Nazar y Cabré, 2012) como los métodos más eficientes en la identificación automática de términos especializados en nuestro corpus legal, pues llegan a alcanzar picos de precisión de 88% y 84.5% respectivamente para los primeros 200 candidatos a término consiguiendo una media de 73% en el caso de *TermoStat* y de 71.5% en el de *Terminus.*

Una vez seleccionados se implementan en *BLaRC,* el corpus de 8,85 millones de palabras, obteniendo similares resultados. Por último, la sección 3.2.4. presenta dos listados validados de 541 términos mono-léxicos (identificados por *TermoStat* sin incluir las coincidencias con el segundo listado) y 2.310 mono-léxicos y poli-léxicos (identificados por *Terminus*) extraídos de este corpus.

Para finalizar este capítulo se lleva a cabo una revisión de la literatura relacionada con el aprendizaje basado en corpus (DDL), continuando con la propuesta de actividades dedicadas a la explotación didáctica de *BLaRC,* nuestro corpus especializado, y de los inventarios de vocabulario técnico obtenidos tras la aplicación de los métodos ATR descritos anteriormente. En ellas se trabajan distintos niveles

lingüísticos tales como el morfológico, sintáctico, semántico o discursivo con el fin de demostrar las posibles aplicaciones de los datos obtenidos de manera automática en secciones anteriores.

El capítulo 4 se dedica al estudio del vocabulario sub-técnico. En el apartado 4.2. se discute el concepto de sub-tecnicidad según las definiciones de diversos autores (Cowan, 1974; Trimble, 1985; Baker, 1988; Farrell 1990; Flowerdew, 2001; Lan, 2001; Chung y Nation, 2003; Wang y Nation, 2004). En general, todos coinciden en el hecho de que el vocabulario sub-técnico es compartido por los ámbitos general y específico y entre disciplinas científicas. Algunos de ellos destacan el carácter polisémico de estas palabras que adquieren un significado especializado en el contexto específico. Muchos de ellos también señalan su importancia en la enseñanza del inglés para fines específicos (ESP) dado su carácter complejo precisamente debido a su ambivalencia.

Sólo Chung y Nation (2003) y Wang y Nation (2004) son más exhaustivos a la hora de delimitar las características semánticas de este tipo de vocabulario. Basándonos en su taxonomía y la observación de las palabras analizadas en la figura 3 y sus contextos, llegamos a una clasificación del vocabulario legal sub-técnico en tres grupos diferenciados:

1) Palabras que denotan un concepto legal compartido por los campos especializado y general que no cambian de significado en el contexto legal cuyo uso es frecuente en ambos ámbitos: *judge, court, tribunal, law, prosecution, jury, legislation, robbery, theft, guilty, solicitor, tribunal.*

2) Palabras de uso frecuente tanto en el ámbito general como en el específico que cambian de significado en el campo legal compartiendo algunos rasgos semánticos con su significado original: *charge, offence, sentence, claim, decision, grounds, complaint, dismiss, evidence, relief, record, trial, battery.*

x

3) Palabras de uso más frecuente en el ámbito especializado que en el general que cambian de significado en el contexto legal adquiriendo un nuevo significado muy diferente o totalmente alejado del general: *appeal, conviction, party, warrant, terms, act.*

La aplicación del modelo de redes léxicas de Williams (2001) al estudio del vocabulario sub-técnico nos proporciona una cantidad de información sobre su contexto de gran valor a la hora de observar la cantidad y frecuencia de los colocados y co-colocados de este tipo de vocabulario. La tabla 1 nos muestra el número y frecuencia de los elementos constituyentes de las redes léxicas de las palabras analizadas en la sección 4.3 tanto en el corpus especializado (*BLaRC*) como en el general (*LACELL*).

Tabla 1. *Cantidad y frecuencia de los colocados en las redes léxicas obtenidas de BLaRC y LACELL.*

Palabra	Colocados y co-colocados en *BLaRC* (normalizado)	Colocados y co-colocados en *LACELL* (normalizado)	Frecuencia normalizada *BLaRC*	Frecuencia normalizada *LACELL*
PURSUANT	404.40	0	10.34	0
ESTOPPEL	114.57	0	8.65	0
LIABILITY	421.69	0	8.20	0
BATTERY	27.57	0.73	7.89	2.27
CONVICTION	281.35	1.33	10.41	3.23
SENTENCE	491.25	1.53	9.50	2.98
DISMISS	338.64	3.20	10.06	3.81
SOLICITOR	159.77	0.33	8.23	2.39
RELIEF	184.18	6.08	9.88	4.45
TRIAL	666.66	2.33	9.22	3.84
LEGISLATION	246.44	39.7	9.23	4.2
WARRANT	30.39	1.60	7.91	3.01
PARTY	708.36	274.13	9.22	4.73
CHARGE	167.68	64.77	9.08	4.89
COMPLAINT	180.22	18.18	8.79	4.70
OFFENCE	522.93	28	8.91	5.03
GUILTY	66.55	11.96	6.87	4.25
EAT	0	2.20	0	3.27
BLUE	0	13.43	0	3.52
MORNING	0	268.36	0	4.94

En la tabla 1 se observa que tanto el número de colocados como la frecuencia de éstos es mayor en el corpus específico en tanto en cuanto el término es más especializado y viceversa. Por este motivo se propone un método cuantitativo −descrito en la sección 4.3.− para establecer un coeficiente de sub-tecnicidad por el que este tipo de palabras puedan localizarse a lo largo de un continuum de especialización en función de la cantidad y la frecuencia de los colocados en cada una de sus redes léxicas especializada y general.

El coeficiente de sub-tecnicidad de una palabra $ST(w_i)$ se calculará restando la frecuencia media de los colocados y co-colocados en el corpus general $\overline{\mu_i^G}$ del mismo parámetro en el corpus especializado $\overline{\mu_i^T}$. Ambos valores se normalizarán dividiédolos por el número de tokens en cada corpus, esto es:

$$ST(w_i) = \frac{\overline{\mu_i^T}}{|C^T|} - \frac{\overline{\mu_i^G}}{|C^G|}$$

Donde $|C^T|$ y $|C^G|$ repesentan el número de tokens en los corpora especializado y general respectivamente. Los resultados de la implementación de este método se muestran en la figura 3.
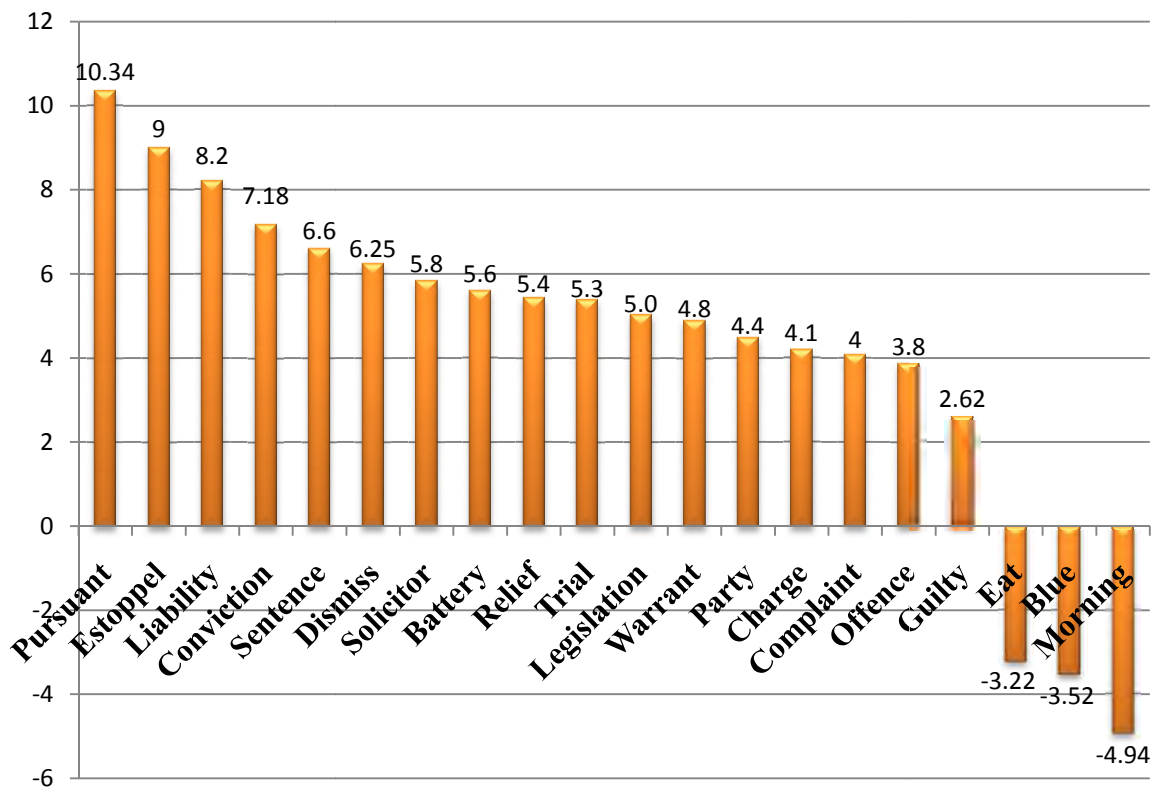
Figure 3. *Coeficiente de sub-tecnicidad*

Sin embargo, la descripción del vocabulario sub-técnico no sería completa sin un examen de carácter semántico de este tipo de palabras. Cantos y Sánchez (2001) ofrecen un modelo de análisis lingüístico, las *constelaciones léxicas,* que nos permite representar de una manera muy visual el proceso por el que las palabras sub-técnicas adquieren nuevos rasgos semánticos que de uno u otro modo se relacionan con su significado original. Las constelaciones léxicas muestran cómo el lenguaje se estructura de una manera jerárquica por la que unas palabras atraen a otras y éstas a su vez a otras diferentes creando una intricada red de relaciones del mismo modo que los planetas orbitan en torno a una estrella y ésta se integra en un sistema mucho más amplio, de ahí su denominación.

La sección 4.4. se dedica al estudio semántico de tres términos sub-técnicos, *trial, charge* y *battery,* siguiendo el método de las constelaciones léxicas de Cantos y Sánchez (2001), ya aplicado al inglés específico de las telecomunicaciones por Rea y Sánchez (2010). La figura 4 representa la constelación léxica de *charge* mostrando cómo el significado legal de esta palabra se relaciona con su significado original, "peso, carga". Un "cargo" entendido como acusación podría considerarse como un peso figurativo que recae en los hombros del acusado de la misma manera que una responsabilidad o una deuda pueden resultar "cargas" que dificulten la vida diaria de aquel que las acarrea. Todos estos rasgos semánticos se añaden al significado original de la palabra *charge* facilitando la adquisición de nuevas acepciones, entre las cuales se encuentra su significado especializado. Aun así, como se puede apreciar en la figura 4, existe una cierta proximidad semántica entre el significado base y el legal no resultando excesivamente complicado establecer una asociación entre el concepto de *charge* como "carga" o "peso" y el de "acusación", como ya se ha indicado con anterioridad.

Este método de análisis léxico proporciona una oportunidad única de visualizar las jerarquías semánticas existentes entre los distintos significados de una misma palabra y cómo éstos interactúan entre sí llevando del ámbito general al específico mediante la adquisición de nuevos rasgos semánticos dependientes de niveles jerárquicos superiores.
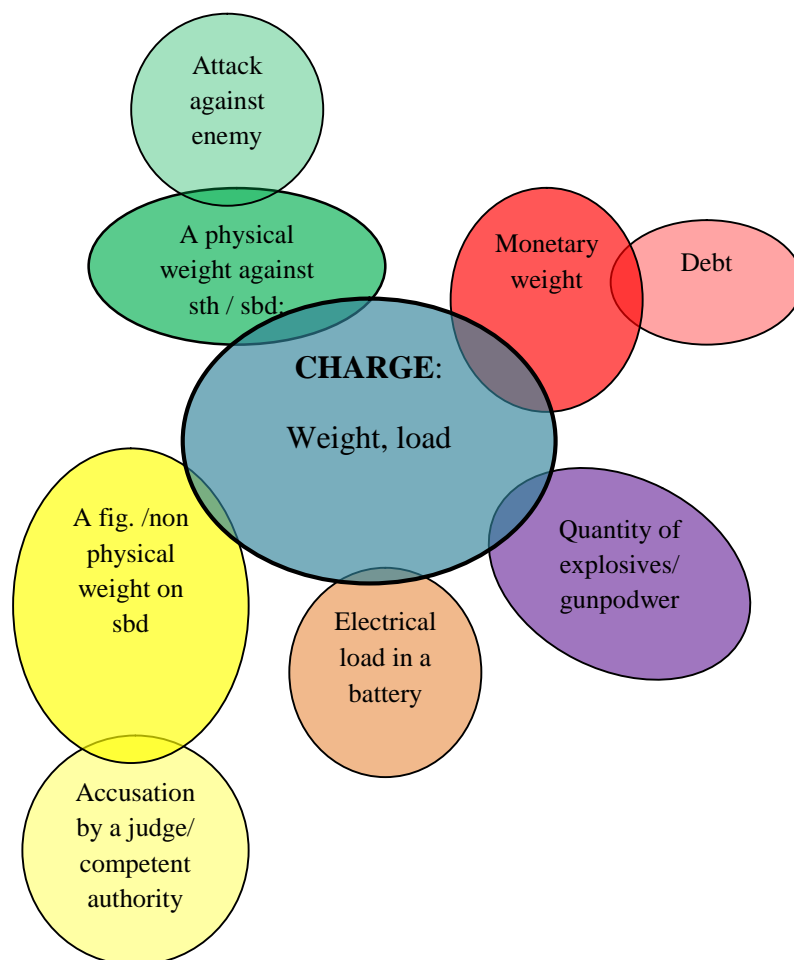
Figura 4. *Constelación léxica de CHARGE*

CONCLUSIÓN Y LIMITACIONES DEL ESTUDIO

Los objetivos alcanzados en esta tesis han sido, por un lado, la identificación del vocabulario especializado de *BLaRC,* un corpus de sentencias judiciales del Reino Unido compilado *ad hoc.* Con ese fin, se han evaluado diez métodos ATR para seleccionar aquellos más eficientes en el reconocimiento automático de los términos legales de nuestro corpus. Para ello se ha utilizado un corpus piloto de 2,6 millones de palabras extraído de *BLaRC* implementando sobre él los diez métodos ATR. Se han

calculado los niveles de precisión media y cumulativa para cada uno de los métodos mediante la comparación con un *gold standard,* identificando *Terminus 2.0* (Nazar y Cabré, 2012) y *TermoStat* (Drouin, 2003) como los más eficientes. Finalmente se han implementado ambos métodos en *BLaRC* elaborando dos listados de 2.851 términos legales mono-léxicos y poli-léxicos validándolos automática y manualmente.

El capítulo 4 se ha dedicado al análisis cuantitativo y semántico del vocabulario legal sub-técnico dada su importancia dentro del inglés jurídico. Para el análisis cuantitativo se ha planteado un algoritmo, *SubTech,* con el que se puede medir el nivel de especialización de este tipo de palabras en función de su contexto de uso especializado y general, la información relativa a este contexto se ha obtenido gracias a la aplicación del modelo de las redes léxicas de Williams (2001).

Por último, en el apartado 4.4, se lleva a cabo un análisis semántico de tres términos sub-técnicos, *trial, charge,* y *battery,* implementando el modelo de análisis lingüístico de Cantos y Sánchez (2001) conocido como las constelaciones léxicas. Utilizando este modelo se consigue representar de manera visual el proceso de especialización por el que este tipo de palabras adquieren nuevos rasgos del significado derivados de su sentido original. Dicho proceso se visualiza de manera multi-dimensional mostrando claramente la jerarquía de significados existente entre las distintas acepciones de la palabra analizada. Asimismo, las *constelaciones léxicas* permiten representar no sólo los niveles de dependencia semántica entre los distintos rasgos de significado sino también la mayor o menor proximidad semántica entre estos en función de la distancia existente respecto del núcleo de la constelación.

La combinación de ambos métodos para la descripción del vocabulario sub-técnico supone un paso hacia delante en el análisis de un fenómeno léxico que, hasta la fecha, no ha sido explorado con suficiente profundidad.

En lo que se refiere a las limitaciones de este estudio, podría argumentarse, en primer lugar, que el corpus debería cubrir más géneros legales para lograr una mayor representatividad en las conclusiones relativas al léxico de esta variedad. Como ya se ha comentado, la cantidad y variedad de géneros legales es muy amplia y conseguir compilar un corpus de estas características, aunque deseable, sería una tarea muy ardua para un trabajo de investigación de estas características.

Por otro lado, como tema para futuros trabajos de investigación, sería interesante ahondar en el estudio cuantitativo de las palabras sub-técnicas aumentando el número de éstas con el fin de poder llegar a conclusiones más sólidas en lo relativo a una posible correlación entre su clasificación y análisis semántico y los valores obtenidos tras la implementación del algoritmo *Sub-Tech*.

CHAPTER 1
INTRODUCTION

CHAPTER 1

1.1. RATIONALE

Understanding the terms in a specialised text, which encapsulate the specialised knowledge in any subject field, can undoubtedly contribute to a better comprehension of the text itself. Therefore, gaining access to these terms can become an advantage for the researcher interested in studying the characteristics and rules governing the lexicon of a particular variety of language.

As commonly agreed by scholars, legal English (also known as *legalese*) is a peculiarly obscure and convoluted variety of English. D. Mellinkoff, one of the first scholars devoted to the study of *legalese,* affirms that *"*the language of the law has a strong tendency to be: wordy; unclear; pompous [and] dull" (Mellinkoff, 1963: 63). The presence of Latin borrowings and Old French phrases, synonyms, archaisms and redundancy, as well as the widespread use of "common words with uncommon meanings" (Mellinkoff, 1963: 11) characterise its lexicon.

Most of the research carried out to date to describe legal English could be considered eminently prescriptive (Mellinkoff, 1963; Alcaraz, 1994, 2001; Tiersma, 1999; Borja, 2000), that is to say, it is either based on the authors' knowledge and intuitions on the subject or on relatively reduced samples of this language variety.

Authors have profusely discussed on the advantages and disadvantages of employing language corpora as a source of information for linguistic analysis (Sinclair, 1991; Sánchez et *al*., 1995, 2010; McEnery & Wilson, 1996; Dudley-Evans & St. John, 1998; Kennedy, 1998; McEnery et *al*., 2006; Tognini-Bonelli, 2001; Meyer, 2002; Gries & Wulff, 2010; Cheng, 2011; etc.). The chomskyan distinction between *competence* and *performance* stands at the very basis of most early criticism against this

discipline. Intuitive examples reflect linguistic competence as they arise from our tacit knowledge of the language. Conversely, those examples taken from corpora reflect performance, that is, "external evidence of language competence and its usage on particular occasions ... performance is a poor mirror of competence" (McEnery & Wilson, 1996: 6), according to chomskyan linguists. Moreover, this kind of examples is often deemed skewed, frequently leading the linguist to erroneous generalisations on the language.

Nonetheless, due to the fast growth of corpora and processing software nowadays, researchers can rapidly access and analyse large amounts of data that could not even be thought of in the 50s and 60s. Tools like *Sketch Engine* (Kilgariff et *al.*, 2003) allow us to search keywords and concordance lines employing as reference such gigantic corpora as *enTenTen12,* of 12,000 million words. This plethora of data clearly refutes the skewedness argument posed by Chomsky as well as it grants the reliability of the conclusions drawn from the observation of the language samples thus obtained although, as Sánchez et *al*. (1995) underline, the degree to which corpus data should be employed as the only source to base the norm on still remains an open question.

Further to that, intuition should go hand in hand with data collection, as remarked by Partington (1998), and aid the researcher, for instance, to discard ungrammatical examples. Similarly, the direct observation of the data can also contribute to the confirmation of hypotheses or *a priori* formulated theories and call our attention to new aspects of the language that would otherwise remain unnoticed.

As regards the study of the lexicon of sub-languages, when it comes to large specialised corpora, the vocabulary inventories that can be extracted from them represent a valuable source of information that could not be accessed otherwise. However, to the best of our knowledge, the amount of written legal corpora is certainly

limited and the access to them, except for a few cases, is neither complete nor free. As a consequence, we engaged into corpus design and decided to create the *British Law Report Corpus* (*BLaRC*), a legal English corpus of law reports that could act as a source of specific vocabulary to resort to for further linguistic analysis.

1.2. GENERAL OBJECTIVES AND STRUCTURE

The major objective of the present PhD thesis, as stated in its title, is the identification and analysis of the specialised vocabulary in *BLaRC,* an English corpus of law reports designed and compiled *ad hoc*.

As shown in section 2.1., the amount and accessibility of legal corpora is reduced, which led into the compilation of a specialised corpus that could serve the main purpose of this work. *BLaRC* was designed following the main standards in corpus linguistics (CL) as reflected in Sánchez et *al*. (1995) and Wynne (2005) for general corpora, and Pearson (1996) and Rea (2010) for specialised ones. A full description and justification of the corpus as well as the main issues related to its compilation are presented in chapter 2.

Owing to the size of this *ad hoc* legal corpus, 8.85 million words, it became essential to select efficient automatic term recognition (ATR) methods that could reach high precision rates in term identification. This is why ten different ATR methods were tested with the aim of choosing the ones which could extract automatically the greatest amount of both single and multi-word legal terms in it. Chapter 3 is therefore devoted to the description and evaluation of these ATR methods implemented on a pilot corpus of 2.6m words extracted from *BLaRC*. After selecting the two most efficient techniques, they were applied to it. The results of such process are illustrated in chapter 3. This chapter also offers the two validated lists of single and multi-word legal terms produced

by the two most efficient methods selected, *TermoStat* (Drouin, 2003) and *Terminus 2.0* (Nazar & Cabré, 2012). In the final section of chapter 3, several activities based on *BLaRC* and the term lists obtained from it are suggested as a way to illustrate some of the varied applications of these specialised vocabulary inventories.

Both chapters 2 and 3 are dedicated to the identification of the legal vocabulary found in *BLaRC* , whereas chapter 4 focuses on the analysis of part of this inventory. In spite of its level of specialisation, legal language is peculiarly intertwined with general English in a way that other varieties are not. Legal terms can appear in news articles, TV programmes or be used in everyday conversation due to the fact that legal issues are part of our everyday life and culture.

Using Heatley and Nation's (1996) software *Range* to compare the list of single-word legal terms found in *BLaRC* with the most frequent words of English, it appears that 40.47% of these terms are included amongst the most frequent 3,000 word families in West's (1953) *General service list* and Coxhead's (2000) *Academic Word List.* The percentage is slightly higher, 45.41%, if compared with the *British National Corpus* thus confirming that almost half of the legal terminology identified in our corpus is shared with general English.

This is precisely why special attention is paid to shared vocabulary in chapter 4 of this thesis. As shown above, one of the peculiarities of the legal lexicon, which contributes to its ambiguity and poses special difficulties in the teaching and learning of English for Specialised Purposes (ESP), is the use of words which acquire new technical meanings when they get in contact with the legal context, the so-called *sub-technical* or *semi-technical* vocabulary (the concept is reviewed in section 4.2). Such words as *conviction, sentence,* or *trial* are considerably frequent in general English acquiring a new specialised sense when they occur in a specialised environment. There

also exist shared words denoting legal concepts which do not change their meaning in the legal field, as shown in the taxonomy offered in section 4.2.

The fact that *sub-technical* vocabulary is shared by general and specialised English makes it specially hard to extract using quantitative methods, since the statistical data associated to this type of words might be misleading when resorting to corpus comparison. In point of fact, no method has been described to date to try and quantify this phenomenon. That is the reason why section 4.3.2. presents *Sub-Tech,* an algorithm which aims at measuring the degree of specialisation of sub-technical words based on the data provided by the application of William's (2001) lexical network model. After calculating the lexical networks for each of the words examined, both in the specialised corpus (*BLaRC*) and a general English one, *LACELL* (*Lingüística Aplicada Computacional, Enseñanza de Lenguas y Lexicografía[1]*), the data obtained are compared and a sub-technicality coefficient is calculated whereby these words can be allocated along a continuum of specialisation depending on the number and frequency of their collocates in both corpora.

Nevertheless, adopting a different perspective for the description of the path followed by sub-technical words towards specialisation was also necessary for a fuller description of this type of vocabulary. Cantos and Sánchez (2001) offer a novel approach to the process by which words "socialise" with each other. Words attract their collocates in the same way as planets orbit around a star producing structures and sub-structures which are organised hierarchically around a central node. They are the *Lexical Constellations* (LCs) which can be applied to the description of the process of specialisation followed by sub-technical words.

---

[1] For more information on the LACELL research group see: http://www.um.es/grupolacell

Section 4.4. illustrates the application of this model to the study of the words *trial, charge* and *battery* in their acquisition of new semantic features when employed in a legal environment and the relationship of these new features with the original meaning of these words in the general field.

Chapter 5 presents the conclusion to this thesis which gathers the most relevant points and findings in it. It also acknowledges its limitations and the further research derived from it. The conclusion is followed by the bibliographic references and the list of the online resources consulted.

Finally, the appendix section is divided into three main blocks. The first one shows the lists of the top 200 legal terms identified by each of the ATR method tested in chapter 3 after being validated. The second block includes samples of the texts comprised in *BLaRC,* our legal corpus. They are intended to represent the most relevant levels of the institutional pyramid which courts and tribunals are organised into within the United Kingdom, namely, the Supreme Court of the UK; the Privy Council; the High Court of Justice of England and Wales; the Northern Ireland Court of Appeal; the Scottish Sheriff Court and the Magistrates' Court of England and Wales. The third block of appendices provides the lexical networks calculated for all the words analysed in section 4.3. obtained from *BLaRC* and *LACELL*. Owing to their size, they are offered in an enclosed CD-ROM.

CHAPTER 2
CORPUS STRUCTURE
AND JUSTIFICATION

CHAPTER 2

This chapter is devoted to the description and justification of *BLaRC*, the *British Law Report Corpus,* and *ad hoc* legal English corpus which has been designed and compiled to attain the major objectives of this thesis, which were established in the introductory chapter. As already stated, *BLaRC* has been designed abiding by the corpus linguistics standards stated by Sánchez et al. (1995) and Wynne (2005) for general corpora and Pearson (1998) and Rea (2010) for specialised ones. Let us then discuss the most relevant issues considered prior to and throughout its design and compilation phase.

2.1. *BLaRC* AND OTHER LEGAL CORPORA

Research into specific corpora availability led to a short list of legal corpora which did not satisfy our needs. The first corpus worth mentioning is *BoLC* (*Bononia Legal Corpus*), since this is probably the most comprehensive legal corpus existing due to its selection of texts from varied genres and topics, and also the closest to *BLaRC* especially regarding the genres it covers. It is a multilingual comparable Italian-English corpus, designed as part of a research project at the University of Bologna where John Sinclair played an important role as a consultant. It aims at "representing the two different legal systems, in particular the differences between the civil law and the common law systems" (as stated on the project website[2]). Its English section, of ca. 50m words, covers several legal genres, namely, UK statutes, law reports and statutory instruments. It can be freely accessed through the internet but not downloaded.

---

[2] http://corpora.dslo.unibo.it/bolc_eng.html

However, the rest of corpora described herein were either too small to act as a normative reference, or inaccessible. As a matter of fact, they focused on aspects of the language which were not relevant for this study or were conceived as parallel corpora with a translational or comparative purpose.

The *JRC-Acquis* Corpus is one of them. It is a multilingual parallel corpus which includes European Union legislative texts affecting all member states in 22 different languages. The English section contains 23,545 texts and 34,588,383 words. It is fully accessible and downloadable.

The *CorTec* corpus is a scientific-technical parallel corpus divided into four sections, one of them deals with commercial law and includes agreements and contracts in English and Brazilian Portuguese. It has 1m words per section and has been developed by the Translation and Terminology Centre of the University of St. Paul, Brazil.

As for the *HOLJ* corpus, it is a monolingual synchronic one comprising 188 judgments of the House of Lords from 2001 to 2003. The number of words is approximately 3,000,000 and its aim is to define a set of rhetorical role labels.

Lastly, the *Cambridge International Corpus*, owned by Cambridge University Press, has a legal corpus section of 20m words. It is neither accessible nor commercialised. It has been employed by CUP to design their legal English books.

There also exist legal sections or materials included in some of the best known general British English corpora like the *BNC* (*British National Corpus*[3]) or the *COBUILD*, but they could not serve our purpose either as they are non-specific and cannot be freely downloaded or processed[4].

---

[3] For more information on the *BNC*: http://www.natcorp.ox.ac.uk
[4] Tom Cobb's website (http://www.lextutor.ca/concordancers/concord_e.html) offers the possibility of freely consulting a 2m word legal section of BNC.

2.2. LAW REPORTS AND THEIR ROLE IN COMMON-LAW-BASED LEGAL SYSTEMS

Establishing the sampling frame, that is, "the entire population of texts from which we [would] take our samples", as McEnery and Wilson (1996: 78) put it, was our first objective, and law reports were selected due to the pivotal role they play in the UK judicial system as well as in any other common law countries. We are aware that the conclusions drawn from the study of one single genre cannot be extrapolated to the whole variety, one of Chomsky's criticisms against early CL. However, law reports, that is to say, written reports of judicial decisions or judgments, stand at the very core of common law systems acting as the main source of law followed by statutes and equity, hence the relevance of focusing on this legal genre and its lexicon.

If representativeness is crucial for the design of any corpus (Sinclair, 1991; Biber, 1993; Sánchez et *al.*, 1995; McEnery & Wilson, 1996; Wynne, 2005, etc.), narrowing the boundaries of our object of study became a must, as we soon realised how legal language is intertwined with everyday language, how it is present both in the public and private fields, and consequently how the vastness of this ESP branch could not be covered or managed in a project of this nature. The following legal genre taxonomy offered by Orts (2006) illustrates this fact:

> **"Public Law:**
> a) Unenacted law:
> - Judgments (that is, the content of law reports)
> - Subpoenas, summons, injunctions
> b) Enacted law:
> - Enactments, statutes,
> - Delegated Legislation
> **Private Law:**

- Wills, deeds, underwritings

- Power of attorneys, divorce agreements

- Contracts (leases, sales contracts, export documents, insurance

policies, arbitration clauses, etc.)

Doctrine and Jurisprudence: textbooks, casebooks, articles,

manuals, etc." (Orts, 2006: 119)

Orts (2006, 2009) offers a comprehensive review of different approaches to *legalese* and legal genres both from the field of law (Melinkoff, 1963; Jackson, 1985; Tiersma, 1999; etc.) and linguistics (Crystal & Davy, 1969; Danet, 1980; Bathia, 1993, 2004; Kurzon, 1986; Swales, 1985; Maley, 1987; Alcaraz 1990, 1994, 2000; etc.). The number of legal genres authors have identified varies depending on the perspective of their analysis, and law reports appear in generic classifications as part of the oral mode (Danet, 1980); within the category "recording and law making" (Maley, 1994); or as public unenacted law (Orts, 2009), amongst others.

Sinclair states that "the contents of the corpus should be selected … according to their communicative function in the community in which they arise" (in Wynne, 2005: 5), therefore, the selection of law reports as the object of study could be justified owing to the fact that they are an essential wheel in the British legal machinery and their status within it is unquestionable.

The United Kingdom belongs to the realm of common law, as opposed to civil or continental law. Western European law, except for the UK, is based on the civil law system. Although it may refer to and apply the existing jurisprudence, it mostly relies upon the law pertaining to the criminal or civil fields (amongst others) which is codified following the Roman law tradition. On the other hand, in common law countries like USA, Canada, Australia, etc., and specifically in the UK, law decisions were based on

previous cases always abiding by the principle of *stare decisis* (to stand by what has previously been decided), and not on acts passed at the parliament.

Nevertheless, common law systems have evolved in different ways: some of them are mixed like Québec or Scotland, where the law is both codified and uncodified. The majority of them is mostly jurisprudential and complies with the principle of binding precedent, that is to say, the decisions made at a higher tribunal should act as binding precedent as long as they are related to the case in question in their essence. Determining what the essence of a given case is –establishing the *ratio dicendi*– is part of the judge's role. "Cases must be decided the same way when their material facts are the same, ... but the legally material facts may recur and it is with these that the doctrine is concerned", according to Williams (in Bhatia, 1993: 128).

Nonetheless, in purely common law systems, the acts passed at their parliaments have gained greater importance being most often cited in case decisions. In the last 150 years (Orts, 2006), enacted law has become essential as a source of law, albeit law reports, as far as they interpret the law and the existing precedents, stand out as the major one.

Another relevant communicative function, as highlighted by Bathia (1993: 119-120), is the one played by this legal genre within Higher Education. The use of law reports as an essential reference for Law students makes them fundamental for this ESP variety. Furthermore, law reports are rather comprehensive texts since they not only cover all the branches of Law, but also touch upon other genres like statutes, wills, contracts, etc. when such text types are referred to as facts, evidence, or any other section within the judicial decision, hence their relevance from a linguistic point of view.

Law reports are written reports of judicial decisions on cases that solicitors, barristers[5], judges, or any other legal professionals need to know. They must be cited and act as the solid ground on which they will build their arguments. This is why, in those common law systems, law reports are made public through different institutions, i.e. the Incorporated Council of Law Reports of England and Wales (ICLR), publishing houses like Butterworth or Lloyds, etc., every year. Due to the widespread use of information technologies and particularly the internet, there is a tendency towards digitalising these texts and storing them in online databases. Using search engines can make case citation a really easy task that used to take ages for legal practitioners to become fully informed about.

There are voices which stand against such availability of case decisions as authorised as The Lord Chief of Justice's[6]. In the launch of the ILCR's DVD on law reporting (October 2009), Lord Judge stated that "all too many cases cited in court, … had simply been downloaded from the internet with no regard to their value as precedents". Whether this be right or not, for a linguist designing a legal corpus, this is an undoubtedly valuable source of information about the way this variety of English is used in real situations.

Nonetheless, access to most of these data bases is often restricted, there are such popular ones as Justis.com, LexisNexis.com, etc. (they are really expensive due to the amount of time they save, so law firms, law faculties, and the like are subscribed users precisely because of that). These data bases offer different possibilities to legal

---

[5] The terminology in use refers solely to England and Wales as it varies considerably from one system to the other one. Solicitors are lawyers who do not have right of audience, they can only draft legal documents but cannot represent their clients at court, this is the function of barristers who can act as counsel for defence or prosecution at most courts (except for tribunals where solicitors are allowed to do it).

[6] He is the head of the judiciary in England and Wales.

practitioners to locate and cite cases depending on the court they were heard at, their main topic, the judges who heard them, the identity of the parties, etc.

However, the British and Irish Legal Information Institute (BAILII.org) has created a completely free and comprehensive online data base with more than 200,000 cases available (about 11 gigabytes of legal materials) and classified them according to the court where they originated and the jurisdiction they belong to.

Although we are not subscribed users of the data bases mentioned above, we have enjoyed free access to some of them and confirmed the fact that, leaving aside the numerous possibilities and applications they provide to legal practitioners −who they were designed for−, they offer a smaller amount of texts than the free-access BAILII database. BAILII has become a really useful and free source of not only case decisions (most of them), but also statutes and some scientific legal texts. It is supported by a number of sponsors like the Inns of Court (barristers' professional associations), law faculties (Cambridge, Oxford, Glasgow, Edinburgh, Cork, etc.), law firms and other prestigious institutions, hence its importance and recognition by professionals.

2.3. GEOGRAPHIC SCOPE OF CORPUS TEXTS

As well as abiding by hierarchical criteria when organizing the corpus, one of the first elements that conditioned our choice was the way that legal vocabulary varies according to the system where it is used. This is so because of the laws and regulations that organise the countries which the UK is divided into. The judicial systems of Northern Ireland, Scotland, England and Wales do not solely depend on UK institutions, but rather have their own autonomous systems and structure. But for the Supreme Court (in general terms) and the UK Tribunal Service (except for some cases), each country is fully independent as regards its judicial system.

This being so, *BLaRC* was structured into five main branches depending on the jurisdictions of their judicial systems, that is, the geographical scope of their courts and tribunals:

1. Commonwealth countries.

2. United Kingdom.

3. England and Wales.

4. Northern Ireland.

5. Scotland.

Special attention is deserved by the first section, that of Commonwealth countries. The Judicial Committee of the Privy Council is a UK institution whose main role is acting as the "highest court of appeal for many current and former Commonwealth countries, as well as the United Kingdom's overseas territories, crown dependencies, and military sovereign base areas" (as stated on their website[7]). The cases heard at this court may come from such varied origins as Mauritius, Caiman Islands, Jamaica, Trinidad and Tobago, etc. Since such geographical variation necessarily implies terminological changes due to their different legal systems, it seemed interesting to devote one of the sections of the corpus to the texts coming from such varied sources, in spite of them not being too numerous.

As regards the second section, it comprises those institutions which are competent to judge cases from all over the UK (with certain exceptions). This category includes the court of last resort of Great Britain, the Supreme Court, as well as the net of administrative courts.

The other three sections are organised in the same way as their judicial systems, that is, except for England and Wales which share the same structure and laws, the

---

[7] http://www.jcpc.gov.uk/

justice of Northern Ireland and Scotland work independently from the other two but for the net of administrative tribunals (barring some cases), and the Supreme Court, as already indicated.

2.4. CHRONOLOGY: DATE OF TEXTS

BLaRC is a specific synchronic monolingual corpus of legal English texts which has been designed and compiled with the aim of identifying and studying the specialised vocabulary of law reports in the United Kingdom. Following Pearson, "a specific corpus compiled for terminological studies, [should include texts] … delivered in the last 10 years prior to the date of compilation" (1998: 51). This is why the texts included in this corpus were produced at UK courts and tribunals from 2008 to 2010. The texts were always gathered randomly yet always belonging to the time span just mentioned.

Moreover, owing to the changes that the structure of these courts has experienced as a consequence of the recent modifications of the law that regulates it, we considered that, if the structure of the corpus responded to the structure of UK courts and tribunals because of thematic and hierarchical reasons −as will be shown below−, it should adjust to the latest modifications it has experimented.

We are specifically referring to the *Constitutional Reform Act, 2005*, by which the Supreme Court of the United Kingdom was created and started to work on 1st October 2009. Its role was formerly performed by the so-called Law Lords of the House of Lords (one of the two chambers the British Parliament is divided into), and the *Tribunals, Courts and Enforcement Act, 2007* which regulates the structure of these institutions thus affecting the structure of *BLaRC* itself.

2.5. MODE AND DOMAIN OF TEXTS

The mode of the texts included in *BLaRC* is written. They were all stored in raw text format using code labels to facilitate their identification. The codes specify the jurisdictional area they belong to, for instance, *EW* for England and Wales or *NI* for Northern Ireland; the court or tribunal, *SC* would stand for the *Supreme Court*; and the order number the texts had been assigned. Thus, if a text was labelled *<EWHCFAM1>*, it would indicate that it is the first one in a category where only the judicial decisions made at the High Court of Justice (*HC*), Family Division (*FAM*) of England and Wales (*EW*) would be included. These codes also facilitate their processing with *Wordsmith, 5.0* (Scott, 2008)*,* the software tool employed to produce the type lists necessary to implement automatic term recognition (ATR) methods.

The exclusion of oral samples of legal language is justified by the difficulty of having access to such material. Obtaining this kind of samples would have implied having access to courtrooms and permission to record the trial sessions, a certainly complicated objective for Spanish researchers merely interested in linguistic data.

Furthermore, supposing we had been granted access and permission to do so, obtaining an amount of texts that could make our conclusions representative of the variety would have taken ages. Moreover, the range of the text selection included in *BLaRC* would have required going to one and every courtroom belonging to all the jurisdictions and levels the corpus has been structured into, a definitely unattainable task for a project of this nature.

Regarding the texts themselves, they are full authentic transcriptions of judicial decisions as produced by the official court shorthand writers whose structure may vary depending on the nature of the case and the hierarchical position of the court where it was heard. That is to say, cases heard at the Supreme Court follow a complex and long

route of appeal that implies much greater argumentation and case citation than a case tried at a first-tier tribunal (at the bottom of the judicial pyramid).

They are entire texts obtained in digital format from BAILII.org, a free online legal database, as explained above. This was certainly a great advantage that saved much time as regards the compilation phase. The texts were automatically downloaded from the internet using a webcrawler software which allows the user to scan websites and save all the files stored in their servers automatically so that everything in them can be consulted offline afterwards. Once the files had been downloaded, they had to be classified into different folders according to the structure of the corpus. They were also manually supervised to avoid problematic characters which may interfere with their processing and they were assigned a code, as explained above.

The size of the texts varies from really long ones (a minority) of 20,000 words, to really brief ones of about 600. The average is 2,000 to 2,500 words. They have all been produced (though not transcribed) by British judges and reflect their decisions about the cases in question as well as the facts, arguments, prior decisions made at other courts and any other kind of information relevant to the case. There is therefore no similarity in terms of text size among each of the linguistic samples that form *BLaRC* as we do follow the recommendations made by Sinclair in this respect: "Samples of language for a corpus should wherever possible consist of entire documents or transcripts of complete speech events" (in Wynne, 2005). Morever, Biber (1998) refers to the inclusion of long texts in a corpus. Although he attested that "the counts are relatively stable across 1,000 word samples ... some grammatical features ... are so rare that they would require larger samples".

19

2.6. CORPUS SIZE AND REPRESENTATIVENESS: ESTABLISHING THE WORD TARGET

Representativeness is central to corpus design and the size of a corpus may determine whether it is representative of the variety of the language it aims at covering or simply an illustrative sample of it with no predictive value. Sánchez et al. (1995) highlight the relevance of the reliability of the conclusions based on linguistic corpora. Their representativeness is directly linked to the internal structure and organisation of the corpus in order for the conclusions based on it to be susceptible of becoming generalisations on the language. Along these lines, Biber also insists on the importance of this issue owing to the fact that "a corpus is not simply a collection of texts. Rather, a corpus seeks to represent a language or some part of a language" (Biber, 1998: 246).

Nonetheless, there seems to be no clear agreement as regards the recommended size for a specialised corpus. Most approaches to this question are made on a theoretical basis. Whereas Pearson (1998) proposes a million words as a reasonable number (she poses that the limit should rather be established by the number of texts available and convertible into digital format), Sinclair (1991) believes that corpora must be as large as possible, establishing 10 to 20m words as the recommendable target for a specialised one. On the other hand, Kennedy (1998) does not consider that a big corpus necessarily represents the language better than a small one. In addition to this, Flowerdale underlines that the size of a specialised corpus will necessarily depend on the aim the corpus has been designed for, given that "specialised corpora are constructed with an *a priori* purpose in mind" (Flowerdale, 2004: 25).

Only a few authors draw their conclusions in this respect from actual data. Heaps (1978), Young-Mi Yeong (1995) or Sánchez and Cantos (1997) propose measures to try and determine the most suitable size for a corpus based on regression

techniques. Nevertheless, as acknowledged by Cantos (Yang et *al*., 2000), these studies present certain limitations since "the functions proposed may be likely to change as the corpus grows dramatically" (Yang et *al*., 2000: 21). Based on the work by Sánchez and Cantos (1997), Corpas (2010) suggests that observing the way lexical density evolves in a corpus as its size augments might be indicative of the ideal size it should reach. Instead of concentrating on the growth of the number of tokens, Corpas relates the evolution of the type/token ratio to the amount of texts included in the corpus, assuming that once a given number of documents has been reached, the number of types does not increase parallel to the number of tokens.

The data offered below are based on Sánchez and Cantos' (1997) proposal to formulate a method to try to determine the optimum size for a corpus to be representative of given language variety. These authors divide the *CUMBRE* corpus, a Spanish text collection of 8 million words, into several mini-corpora of similar size, also respecting the reference corpus internal structure, with the aim of designing a formula that can predict the number of types and lemmas in relation to the number of tokens. Thus, researchers aiming at compiling a corpus could save time in gathering an excessive amount of data by applying these formulas to a relatively small amount of texts.

Sánchez and Cantos (1997) demonstrate that, while the number of tokens augments linearly, the number of types and lemmas is represented by a parabolic function. Several tests are carried out to confirm the validity of the formulas they propose showing that the *TYT-formula* manages to predict the number of types and lemmas in a corpus with a ±5% error margin, "and this speaks eloquently of its validity" (Sánchez & Cantos, 1997: 276).

Therefore, taking all these perspectives into consideration, it became necessary to confirm that our initial word target (8.85m words) would suffice for a study of this kind, which focuses on term identification and analysis. In order to do that, following Sánchez and Cantos (1997), the type/token ratio in *BLaRC* was measured as a potentially good indicator of its lexical density. This was done using *Wordsmith 5.0* (Scott, 2008), which allows us to calculate this ratio automatically. The corpus was divided into 27 sub-corpora of similar size (ca. 2Mb each) which were progressively brought together so that the number of tokens augmented homogenously. These sub-corpora were organised and structured respecting the thematic areas the texts belonged to (as long as it was possible) in order to grant the reliability of the results obtained. Then, the sub-corpora were processed to observe how the type/token ratio evolved as the amount of running words increased.

Table 1.

Standardised type/token ratio and type increase in BLaRC

TOKENS	TOKEN INCREASE	TTR	TYPES	TYPE INCREASE
297,097	100%	33.88	10,271	100%
561,454	47.08%	33.95	14,272	28.03%
985,797	43.04%	33.86	18,397	22.42%
1,249,732	21.11%	33.95	20,940	12.14%
1,570,292	20.41%	34.09	23,570	11.15%
1,934,321	18.81%	34.11	25,448	7.37%
2,265,164	14.60%	34.15	27,419	7.18%
2,602,152	12.95%	34.13	29,673	7.59%
2,983,539	12.78%	34.15	31,946	7.11%
3,375,415	11.60%	34.21	34,282	6.81%
3,674,688	8.14%	34.22	35,877	4.44%
3,967,404	7.37%	34.2	37,295	3.80%
4,275,997	7.21%	34.05	38,475	3.06%
4,561,296	6.25%	33.99	39,929	3.64%
4,977,097	8.35%	33.93	41,614	4.04%
5,200,469	4.29%	33.91	42,394	1.83%

5,608,559	7.27%	33.92	43,671	2.92%
5,891,288	4.79%	33.84	44,658	2.21%
6,195,447	4.90%	33.93	45,659	2.19%
6,497,335	4.64%	34.02	46,877	2.59%
6,784,383	4.23%	34.04	47,726	1.77%
7,038,765	3.61%	34.07	48,359	1.30%
7,222,673	2.54%	34.05	49,144	1.59%
7,534,814	4.14%	34.08	50,065	1.84%
7,849,406	4.01%	34.12	50,999	1.83%
8,198,039	3.62%	34.11	52,016	1.78%
8,508,883	3.65%	34.13	53,083	2.01%
8,857,197	3.72%	34.08	53,714	1.17%

Table 1 illustrates how the standardised type/token ratio (computed every 1,000 tokens) does not vary significantly remaining constant from the smallest to the biggest token set. There is a difference of 0.38 points between the lowest and the highest type/token ratio value which may imply that lexical density is very similar regardless of the size of the sub-corpora included in each section.

However, calculating the increase in the number of types with respect to the number of tokens provides more relevant data which could be employed to confirm that the amount of texts collected may suffice to achieve the goals established for this analysis.

As a matter of fact, judging by the progression of the number of types, it can be observed that it is inversely proportional to corpus size as already proved by Sánchez and Cantos (1997). The highest percentages of type increase can be found between ca. 300,000 and 1.5 million tokens decreasing progressively from 28.03% to 11.15%. Once the corpus reaches 3.6 million words (tokens), type increase drops considerably from 6.81% to 4.44%. This percentage diminishes to less than 2% when the corpus doubles its size from 3.5 to 7 million tokens fluctuating slightly as it continues growing. Finally,

it is hardly significant once the corpus augments to 8.3 million tokens falling from 2% to 1.17% (8.83m).

Additionally to Sánchez and Cantos' (1997) study, term increase was also measured following the same procedure as the one suggested by these authors. The terms in the type list were identified by comparison with a specialised legal English glossary of 10,088 terms described in detail in chapter 3. Both the glossary and the lists generated by *Wordsmith* (after progressively bringing together the 27 sub-corpora *BLaRC* was divided into) were compared using an excel spreadsheet so as to find out how many true terms (TTs) were included in each of the lists of types obtained.

The graph in figure 1 illustrates how the percentage of type and term increase follows a very similar fashion revealing that the latter is also inversely proportional to the number of tokens displayed on the x-axis. The first set of tokens includes sub-corpora 1 and 2, the second one, sub-corpora 1 to 3, the third one, sub-corpora 1 to 4, and so forth.

Concerning the number of new terms appearing as the corpus grows bigger, the graph indicates that once the corpus reaches 1 million tokens, it augments considerably less falling from 10.03% to 4.72%. From that point on, although slightly recovering, this percentage will drop to 1.62% for sub-corpora 1 to 7 (2.26m tokens). It remains constant at 1.02% on average until the corpus grows to 6.78 million words decreasing to 0.4% and not experimenting any significant changes from that point on.
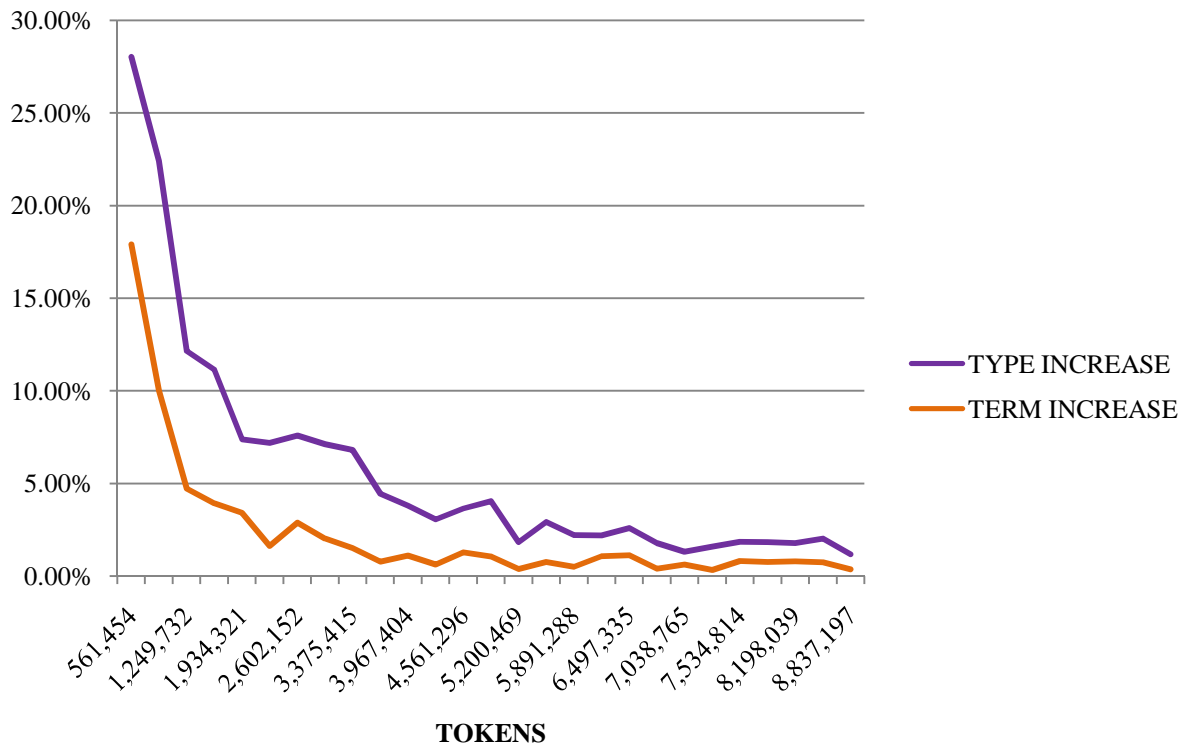
Figure 1. *Type/term increase in BlaRC*

Consequently, it appears that the initial target established for a corpus like *BLaRC* may suffice to attain the objectives set for its compilation. In point of fact, 3.5 million words would have been enough due to the low increase in the percentage of new types and terms appearing as the corpus grew bigger. This is the reason why a pilot corpus of 2.6 million words (*The United Kingdom Supreme Court Corpus)* was extracted from *BLaRC* in order to facilitate the process followed to validate the automatic term recognition methods examined in chapter 3.

In spite of that fact and owing to the major aim of this work, that of identifying and analysing the specialised vocabulary in our legal corpus, letting the corpus grow to 8.85m words would necessarily cause the number of terms to augment (however low such increase might appear to be), providing more detailed information on the specialised lexicon of the corpus and facilitating such tasks as, for instance, the analysis

of rarer cases such as highly technical terms presenting lower frequency and distribution values.

2.7. LEXICAL VARIETY

Law reports should not only be paid special attention within ESP because of their essential function in common law systems, but also because of their vast topic coverage. This corpus has been organised according to the source where the corpus texts originated, that is, what court or tribunal cases were heard at and decided on.

Tribunals and courts are specialized in a given branch of law: criminal law, family law, commercial law, intellectual law, etc., and law reports touch upon one and every branch of both the private and public fields. Judges are in charge of judging cases by both interpreting the law itself (the statutes passed at the parliament), and fundamentally taking into consideration the existing precedents. Therefore their judgments, as reflected on law reports, pertain to all the fields of law.

2.8. THE STRUCTURE OF COURTS AND TRIBUNALS AND THEIR PARALLELISM WITH *BLaRC*

In *Corpus Linguistics,* McEnery and Wilson (1996) refer to Biber when highlighting the importance of establishing a clear structure for the design of a corpus prior to its compilation and analysis: "Biber … emphasises the advantage of determining beforehand the hierarchical structure (or strata) of the population, that is, defining what different genres, channels and so on it is made up of" (McEnery and Wilson, 1996:79).

This is the reason why it is essential to justify the categorization method that was followed in the organisation of *BLaRC*. This corpus retains the current UK tribunal and court structure after its recent modifications as reflected on BAILII, the online legal

database which has served as the source to obtain the texts that make it up. This is so because of several reasons, the first one being the relevance of the hierarchy of courts and tribunals in the UK legal system. The principle of binding precedent, which the British judicial system revolves around, establishes that any decision made at a higher court or tribunal will set binding precedent as long as the case is similar to the one under examination in its essence (the *ratio dicendi*).

Secondly, if this structure was maintained, the texts would be grouped according to the field of law they belong to, so they would be similar in lexical terms, and comparing results by studying the categories separately would be easier and respond to a thematic criterion which is fundamental as far as the identification and study of the specialised vocabulary of law reports is concerned.

In the third place, the route of appeal for a case also responds to this hierarchy. One single case could be heard at more than one tribunal or court if it obtained leave of appeal, that is to say, when a decision is not favourable to any of the parties involved in a trial, it may be appealed to and, if granted permission, it could be heard at higher instances. This fact implies that there are similar tribunals and courts belonging to the same field of law at different levels of the judicial structure, i.e. the UK Upper tribunal of Finance and Tax and the First Tier Tax Tribunal deal with similar cases, yet the former is at a higher level and would either have jurisdiction over certain cases which imply, say, greater amounts of evaded money, or others that come from First Tier tribunals and have been granted leave of appeal.

The same case could go up the structure to the court of last resort of the United Kingdom: the Supreme Court, although, as far as the lexical content of the texts is concerned, it should be modified and argued in greater depth every time it is reviewed

and heard at a higher level, thus becoming a different text as it follows the route of appeal.

Another factor conditioning the structure of *BLaRC* is the distribution of the population in the United Kingdom. As it is shown in the UK official census 2011, elaborated by the Office of National Statistics, it appears that almost 90% of the population of the whole territory is concentrated in England and Wales while Northern Ireland only has about 3 % and Scotland 9%. Although the number of texts and word targets per category and subcategory were not distributed mathematically depending on these figures, they were taken into account in order to reinforce the representativeness of the texts obtained from English and Welsh sources that amounted to approximately 55% of the total.

Having justified the way the corpus has been structured, its categorisation goes as follows:

1. **Commonwealth countries**

    1.1.   The judicial committee of the privy council

2. **UK courts and tribunals**

    2.1.   United Kingdom House of Lords and UK Supreme Court.

    2.2.   Upper Tribunal (Administrative Appeals Chamber)

    2.3.   Upper Tribunal (Tax and Chancery)

    2.4.   Upper Tribunal (Immigration and Asylum Chamber)

    2.5.   Upper Tribunal (Lands Chamber)

    2.6.   First-Tier tribunal General Regulatory Chamber.

    2.7.   First-tier Tribunal (Health Education and Social Care Chamber)

    2.8.   First-tier Tribunal (Tax)

    2.9.   United Kingdom Competition Appeal Tribunal

2.10.   United Kingdom Employment Appeal Tribunal

2.11.   United Kingdom Financial Services and Markets Tribunals

2.12.   United Kingdom Asylum and Immigration Tribunal

2.13.   United Kingdom Information Tribunal including the National

Security Appeals Panel

    2.13.1 United Kingdom VAT & Duties Tribunals (Customs)

    2.13.2 United Kingdom VAT & Duties Tribunals (Excise)

    2.13.3. United Kingdom VAT & Duties Tribunals (Insurance

    Premium Tax)

    2.13.4. United Kingdom VAT & Duties Tribunals (Landfill Tax)

2.14.   United Kingdom VAT & Duties Tribunals

2.15.   Nominet UK Dispute Resolution Service

2.16.   Special Immigrations Appeals Commission

2.17.   United Kingdom Special Commissioners of Income Tax

2.18.   UK Social Security and Child Support Commissioners

3.   England and Wales

3.1.   England and Wales Court of Appeal (Civil Division)

3.2.   England and Wales Court of Appeal (Criminal Division)

3.3.   England and Wales High Court (Administrative Court)

3.4.   England and Wales High Court (Admiralty Division)

3.5.   England and Wales High Court (Chancery Division)

3.6.   England and Wales High Court (Commercial Court)

3.7.   England and Wales High Court (Court of Protection)

3.8.   England and Wales High Court (Senior Court Costs Office)

3.9.   England and Wales High Court (Family Division)

2.9. DISTRIBUTIONAL CRITERIA AND TARGETS PER CATEGORY

The amount of texts forming *BLaRC* is not evenly distributed amongst its categories. Great variation was found depending on the text source (court or tribunal). Whereas there were sections where the overall number of texts was remarkably high −the Administrative Chamber of England and Wales high Court section of BAILII offered 1922 cases between 2008 and 2010−, there were others like the United Kingdom VAT & Duties Tribunal which were also exceptionally low with 1 or 2 at most, yet none of these cases represents the average.

The reasons for the irregular distribution of the texts available are varied, in some cases, especially regarding tribunals, they have either started working recently or disappeared due to the *Tribunals, Courts and Enforcement Act, 2007*, cited above. In some others, the high figures coincide with a densely populated area (one of the criteria supporting text distribution within the corpus) or with a court that, due to its high status in the hierarchy (i.e. any of the chambers of the High Court of Justice of England and Wales), is in charge of hearing a very high number of cases.

In addition to this, it is assumed (as there is no physical evidence of it) that the fact that a court or tribunal is less productive in terms of text availability, implies that there are less cases being heard at it. Whether this be certain or not, it is beyond our knowledge since BAILII is a free online database supported by authoritative institutions and built up by reputed contributors who altruistically donate the transcriptions of judge's decisions obtained from official sources.

Nonetheless, and regardless of the possible explanation that may have been found for this phenomenon, the targets established for the sections and subsections of the corpus were kept proportional to the total number of texts available within the covered time span. Therefore, the sub-targets were set according to this criterion: if the

number of texts in a section was higher, they were assigned a larger word target, thus being more representative of the language as that is the proportion they keep in real life, or at least it is assumed to be so.

All the same, a corpus should not be intended to systematise reality in a mathematical way, in this case, we simply intended to be as coherent as possible in every step we took towards corpus design. As Sinclair (2005) puts it when dealing with the issue of sampling a corpus and the structural criteria to employ when designing it: "real life is rarely as tidy as this model suggests" (Sinclair, 2005: 3). Moreover, "We remain ... aware that the corpus may not capture all the patterns of the language, not represent them in precisely the correct proportions. In fact there are no such things as "correct proportions" of components of an unlimited population".

Having said that, the final targets do not coincide mathematically with the ones planned in advance. This is basically due to the fact that the exact extension of the texts could not be controlled as they were gathered randomly so most sub-targets slightly exceed the initial figures although never significantly.

There is just one exception to this proportional distribution of texts and targets. It is the category corresponding to the Supreme Court of the UK. In this case, due to the fact that it is the court of last resort for the whole of the UK which hears appeals from all jurisdictions and areas of law, we considered that it should be given greater relevance precisely because of that. This is why all the available texts from 2008 to 2010 from these two courts were sampled regardless of their proportion with respect to the total amount of texts existing within this time frame.

Having said so, the total number of texts available between 2008 and 2010 was 16,612. Therefore, the targets were established with respect to it, as already stated. The tables below illustrate how this distribution was organised by showing the total number

of texts available per sub-category, their percentage with respect to the total amount of

texts, and the corresponding word target achieved following this proportion.

1. **COMMONWEALTH COUNTRIES**

COURT	AVAILABLE TEXTS	% OF TOTAL	FINAL WORD TARGET
Privy Council	152	0,92%	55,693

2. **UK COURTS AND TRIBUNALS**

COURT/ TRIBUNAL	AVAILABLE TEXTS	% OF TOTAL	FINAL WORD TARGET
Supreme Court	117	0,70%	1,047,260
House of Lords	74	0,44	1,581,655
Upper Tribunal (Administrative Appeals Chamber)	550	3,31%	250,212
Upper Tribunal (Tax and Chancery)	44	0,27%	9,138
Upper Tribunal (Immigration and Asylum Chamber)	59	0,36%	21,866
Upper Tribunal (Lands Chamber)	135	0,82%	69,904
First Tier General Regulatory Chamber	124	0,75%	47,311
First-tier Tribunal (Health Education and Social Care Chamber)	139	0,84%	67,797
First-tier Tribunal (Tax)	865	5,21%	328,012
Competition Appeals Tribunal	100	0,61%	39,795
Nominet UK Dispute Resolution Service	370	2,23%	140,938
Special Immigrations Appeals Commission	24	0,15%	10,195
Employment Appeal Tribunal	971	5,85%	320,216
Financial Services and Markets Tribunal	16	0,1%	9,025
Asylum and Immigration Tribunal	141	0,85%	58,208
Information Tribunal including the National	130	0,79%	53,117

Security Appeals Panel			
Special Commissioners of Income Tax	80	0,49%	36,356
Social Security and Child Support Commissioners	219	1,32%	83,040
VAT & Duties Tribunals (Customs)	20	0,12%	11,479
VAT & Duties Tribunals (Excise)	92	0,56%	34,896
VAT & Duties Tribunals (Insurance Premium Tax)	1	0,01%	7,079
VAT & Duties Tribunals (Landfill Tax)	2	0,02%	9,466
TOTAL			4,246,965

3. ENGLAND AND WALES COURTS AND TRIBUNALS

COURT/ TRIBUNAL	AVAILABLE TEXTS	% OF TOTAL	FINAL WORD TARGET
England and Wales Court of Appeal (Civil Division)	2640	15,89%	956,398
England and Wales Court of Appeal (Criminal Division)	1136	6,84%	414,683
England and Wales High Court (Administrative Court)	2039	12,27%	731,693
England and Wales High Court (Admiralty Division)	17	0,11%	8,842
England and Wales High Court (Chancery Division)	1009	6,07%	366,298
England and Wales High Court (Commercial Court)	379	2,28%	142,701
England and Wales High Court (Court of Protection)	26	0,16%	34,007
England and Wales High Court (Senior Costs Off.)	70	0,43%	29,302
England and Wales High Court (Family Division)	199	1,20%	84,557
England and Wales High Court (Mercantile Court)	8	0,05%	6,152
England and Wales High Court (Patents Court)	105	0,64%	40,420
England and Wales High Court (Queen's Bench Division)	709	4,27%	255,301
England and Wales High	284	1,71%	101,066

Court (Technology and Construction Court)			
England and Wales Patents County Court	12	0,08%	15,242
England and Wales Magistrates' Court (Family)	98	0,59%	33,680
England and Wales County Court (Family)	56	0,34%	20,702
England and Wales Care Standards Tribunal	70	0,43%	27,762
England and Wales Lands Tribunal	115	0,7%	44,004
TOTAL			3,322,810

4. NORTHERN IRELAND COURTS AND TRIBUNALS

COURT/ TRIBUNAL	AVAILABLE TEXTS	% OF TOTAL	FINAL WORD TARGET
Court of Appeal in Northern Ireland	149	0,9%	57,309
Crown Court for Northern Ireland	149	0,9%	55,792
High Court of Justice in Northern Ireland Chancery Division	44	0,27%	13,748
High Court of Justice in Northern Ireland Family Division	53	0,32%	20,435
High Court of Justice in Northern Ireland Queen's Bench Division	470	2,83%	163,081
High Court of Justice in Northern Ireland Master's decisions	27	0,17%	11,338
Fair Employment Tribunal	81	0,49%	30,484
Industrial Tribunals of Northern Ireland	891	5.36%	327,626
Northern Ireland - Social Security and Child Support Commissioners	142	0,86%	56,450
TOTAL			736,263

5. SCOTLAND COURTS AND TRIBUNALS

COURT/ TRIBUNAL	AVAILABLE TEXTS	% OF TOTAL	FINAL WORD TARGET
Scottish Court of Session	794	4,78%	116,351
Scottish High Court of Justiciary	315	1,90%	115,755
Scottish Sheriff Court	100	0,61%	263,360
TOTAL			495,466

After having calculated the word targets per section and subsection, the overall corpus size is 8,857,197 words.

2.10. CONCLUSION

This chapter gives a full description and justification of *BLaRC*, the legal corpus compiled *ad hoc* as a tool to access and study its specialised and sub-technical vocabulary. The compilation of the corpus was motivated by the scarceness of the legal corpora available, as illustrated in section 2.1.

Its design and compilation process was carried out according to CL standards as defined in Sánchez et *al.* (1995) and Wynne (2005) for general corpora and Pearson (1998) and Rea (2010) for specialised ones, so that the results obtained from its subsequent analysis could be worthy, reliable, and useful. An attempt was therefore made to adhere, as far as possible, to both the criteria proposed for text selection and the guidelines for the compilation of specialized corpora from the literature available.

A well-designed corpus creates an excellent opportunity to look into language evidence and perform quantitative and qualitative analyses. *BLaRC* is structured in such a way that it will allow multiple contrastive analyses in relation to the different

parameters governing in its projection, namely, topic variety, types of courts and tribunals, and geographical scope.

CHAPTER 3
EVALUATION OF AUTOMATIC
TERM RECOGNITION METHODS

CHAPTER 3

As already stated in the introductory chapter, understanding the terms in a specialised text facilitates greatly the comprehension of the text itself, since terms could be regarded as conceptual vehicles which can be employed to transmit specialised knowledge amongst scientists, researchers, or professionals in all specialised areas. Nowadays, thanks to the easy access and availability of online information, corpora, even specialised ones, have grown bigger and bigger requiring the implementation of automatic term recognition (ATR) methods for the automatic mining of terms, a task that could not have been performed on a corpus like *BLaRC* manually, due to its size.

This is precisely why, once our legal corpus was designed and compiled, in order to study the specialised lexicon of the legal genre it was based on, law reports, it became necessary to identify efficient ATR methods that could automatically recognise legal terms with as much precision as possible. Therefore, this chapter presents the implementation and validation of ten different single and multi-word term recognition methods tested on *UKSCC,* the *United Kingdom Supreme Court Corpus,* a pilot corpus of 2.6m words extracted from *BLaRC.*

These methods have been grouped and evaluated separately basically due to the precision levels recorded by their authors who, generally speaking, demonstrate that single-word term (SWT) recognition methods are more efficient than those which can extract solely multi-word terms (MWT) or both type of terminological units. Therefore the most effective techniques in term identification will be selected for their later implementation on BLaRC

.

3.1. SINGLE-WORD TERM RECOGNITION METHODS

3.1.1. Introduction

According to scholars, terminology is used to share domain-specific information amongst the members of a specialised community (Rea, 2008). As Kit and Liu put it, "terms are linguistic representations of domain-specific key concepts in a subject field that crystallise our expert knowledge in that subject" (Kit & Liu, 2008: 204), in other words, a term is "a textual realisation of a specialised concept" (Spasic et *al*., 2005: 240). To Chung (2003a: 221-2), terms display distinctive features both qualitatively (e.g. their morphological structure; their meaning) and quantitatively (e.g. their frequency of occurrence). Hence, identifying and extracting the terms in a specialised corpus becomes an essential task when using it as a source of information for further linguistic analysis. However, handling and processing large amounts of data is a time-consuming task and the application of effective ATR methods is essential for the terminologist to draw reliable conclusions on the information retrieved by such methods.

This section is devoted to the evaluation of five SWT recognition methods tested on a 2.6 million-word pilot corpus, *UKSCC*, to facilitate the implementation of the methods and their validation process, whose structure and features will be defined below.

3.1.2. The relevance of SWTs in term identification

ATR methods typically concentrate on MWTs exploring the concepts of *termhood* and *unithood* from different perspectives. Nagakawa and Mori (2002: 1) define *termhood* as "the degree that a linguistic unit is related to a domain-specific concept". According to Kit and Liu (2008: 205), *unithood* establishes "how likely a candidate is to be an atomic

linguistic unit". Nevertheless, these authors consider that *unithood* only serves as a way of discarding those units not displaying a high level of cohesion amongst their possible constituents but does not provide any information about their degree of specificity.

In the past, the literature on ATR methods and software tools has been profusely reviewed (Estopa, 1999; Maynard & Ananiadou 2000; Cabré et *al*. 2001; Drouin 2003; Lemay et *al*. 2005; Pazienza et *al*. 2005; Chung 2003a, 2003b; Kit & Liu 2008 or Vivaldi et *al*. 2012, to name but a few) often classifying them according to the type of information used to extract candidate terms (CT) automatically. Some of the reviewed methods resort to statistical information, amongst them: Church and Hanks (1990), Ahmad et *al*. (1994), Nakagawa and Mori (2002), Chung (2003a), Fahmi et *al*. (2007), Scott (2008) or Kit and Liu (2008). Other authors like Ananiadou (1988), David and Plante (1990), Bourigault (1992) or Dagan and Church (1994) focus on linguistic aspects. The so-called *hybrid methods* rely on both. The work of Justeson and Katz (1995), Daille (1996), Frantzi and Ananiadou (1996; 1999), Jaquemin (2001), Drouin (2003), Barrón Cedeño et *al*. (2009) or Loginova et *al*. (2012) illustrate this trend. As stated by Vivaldi et *al*. (2012), only a few of these methods resort to semantic knowledge, namely, TRUCKS (Maynard & Ananiadou 2000), YATE (Vivaldi, 2001) and MetaMap (Arson and Lang, 2010).

However, the literature on the evaluation of these methods is not so abundant. There are initiatives for the evaluation of ATR methods like the one organised by the Quaero program (Mondary et *al*., 2012) which aims at studying the influence of corpus size and type on the results obtained by these methods as well as the way different versions of the same ATR methods have evolved. Some authors also show their concern about the lack of a standard for ATR evaluation which is often carried out manually or employing a list of terms, a gold standard, which is not systematically described

(Bernier-Colborne, 2012: 1). Some researchers like Sauron, Vivaldi and Rodríguez, or Nazarenko and Zargayouna (in Bernier-Colborne, 2012) have worked on this area although there is still much to be done in this respect.

In spite of the large number of ATR methods existing to date, very few concentrate solely on SWTs, which are neglected to a certain extent assuming that they are easily identifiable specially due to the fact that such parameters as unithood do not need to be considered. Nevertheless, as remarked by Lemay et *al.* (2005), ignoring SWTs implies taking for granted that most specialised terms are multi-word units. Nakagawa and Mori emphasise this idea by giving concrete data on the percentage of MWTs in specific domains: "The majority of domain specific terms are compound nouns, in other words, uninterrupted collocations. 85% of domain specific terms are said to be compound nouns" (Nagakawa & Mori, 2002: 1), yet they do not provide any kind of evidence to support this piece of data.

Owing to that fact, we decided to calculate the actual amount of SWTs in our legal glossary (used as gold standard for comparison), which was compiled after merging and filtering four different electronic legal glossaries[8], finding that STWs represented a much larger amount of terms than Nagakawa and Mori affirm. As a matter of fact, having examined it thoroughly, 65.22% of 10,088 terms in the list were found to be SWTs. Thus, the evaluation of the methods presented below will include the four main lexical categories of the language, namely, nouns, verbs, adjectives and adverbs.

---

[8] Both British and American English legal terms have been included in the glossary although British English predominates. The inclusion of American English obeys to the observation of the texts before starting any evaluation procedure. Some of the texts, due to the nature of the claim, appeal, etc., included American terminology. As a matter of fact, although there are obvious differences, both BrE and AmE have many legal terms in common as shown in specialised dictionaries and glossaries.

3.1.3. *The United Kingdom Supreme Court Corpus* (*UKSCC*): **the pilot corpus**

UKSCC is a 2.6 million-word specialised corpus subset of a larger one: *BLaRC*. It was extracted from it in order to validate the methods described below owing to the size of *BLaRC,* which would have made such processes as the lemmatisation or the manual supervision of the validated lists an unattainable task.

In the light of the data provided in section 2.6, the size of *UKSCC* appears to suffice for the validation of the ATR methods selected given the sharp decrease in the percentage of types and terms once the corpus reaches 2.6 million words. Figure 1 reveals that the number of types and terms augments in inverse proportion to the size of the corpus following a very similar fashion, as already demonstrated. While type increase falls 6 points from 28% to 22% once the corpus reaches 1 million words, the difference is slightly bigger with respect to terms falling from 17.90% to 10% within the same range. As the number of tokens in the corpus grows, the gap becomes greater. Once the corpus expands to 2.6 million tokens, type increase falls 3.4 points, remaining at 7.59%, while term increase stands at 2.8%, 1.5 points less. Even so, given the reduced number of new types and terms appearing from that point on, it appears that *UKSCC* may be large enough to act as a pilot corpus to test ATR methods on.

The Supreme Court was selected as the text source for the pilot corpus because of its relevance within the British judicial system (all the decisions made at the Supreme Court set precedent and are cited whenever applicable), and the wide lexical variety of the documents coming from it. It is at the top of the UK judicial pyramid and deals with cases belonging to all branches of law.

As for its structure, *UKSCC* is a synchronic, monolingual and specialised collection of 193 law reports from the UK Supreme Court and the House of Lords

issued between 2008 and 2010. The documents included in *UKSCC* are authentic judgments as reported by British courts in raw text format.

3.1.4. Description of the methods selected for evaluation

3.1.4.1. Keywords (Scott, 2008)

The *Keywords* tool included in the software package *Wordsmith 5* by Scott (2008) could not be considered as an ATR method *per se*, however, as testing will show below, it can be used as such and it does perform more accurately than others designed specifically to that end. It was chosen due to its popularity and capacity to easily process large amounts of text data providing information on a word's "importance as a content descriptor", in Biber's words (in Gabrielatos, 2011: 5), that is to say, on its keyness. According to Scott (2008b: 184), 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".

Scott's tool was configured to apply Dunning's (1993) log-likelihood calculation (it can also employ the chi-square test to produce a keyword list) since it is a recommended option for long texts such as the ones included in *UKSCC*. For the system to calculate a word's keyness it is necessary to resort to a reference corpus in order to compare it with the specialised one whose keywords we wish to extract. The reference corpus we employed for corpus comparison to implement *Keywords* and Chung's method is *LACELL*, a 21 million-word general English corpus compiled by the *LACELL* research team at the University of Murcia comprising mainly texts from the 1990s. It is a balanced synchronic corpus of general English including both written texts from diverse sources such as newspapers, books (academic, fiction, etc.), magazines, brochures, letters and so forth, and also oral language samples from conversation at

different social levels and registers, debates and group discussions, TV and radio recordings, phone conversations, everyday life situations, classroom talk, etc. Its geographical scope ranges from USA, to Canada, UK and Ireland, however, those texts not coming from the United Kingdom were removed to avoid skewedness in the results reducing the original size to 14.9 million words.

The *BNC* lemmatised lists provided online by Kilgariff[9] were employed as background for reference to implement Kit and Liu's method owing to the fact that both the SC (study corpus) and RC (reference corpus) had to be lemmatised[10]. Therefore, *UKSCC* was also lemmatised using Schmid's (1995) *Tree Tagger[11]* to apply the calculations on lemmata, not on word types[12].

3.1.4.2. TermoStat *(Drouin, 2003)*

Drouin designs *TermoStat*, a free online software[13] for automatic term extraction in French, English, Spanish, Italian and Portuguese which can process raw text files up to 30 Mb. He employs a hybrid technique to detect both single and multi-word CTs and rank them according to their level of specialisation. Its main aim is to reduce the amount of noise produced by other automatic methods by cutting down on the number of items included in the lists generated by the system. With this purpose, the author establishes a test-value threshold of +3.09 "which means that probability of finding the observed frequency is less than 1/1000" (Drouin, 2003: 101) acting as a cut-off point between terms and non-terms.

---

[9] At: http://www.kilgarriff.co.uk/BNClists/lemma.num
[10] The process of lemmatisation consists in retrieving a word's lemma, that is, the root word which other possible realisations of it derive from (e.g. *make* would be the lemma for *made*, *makes*, *making*, etc.). Lemma frequency must be computed by adding up the raw frequency values of all its posible variants.
[11] Available at: http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger
[12] The term *word type* refers to every different word form in the corpus but not to each of its occurrences known as *tokens*.
[13] Available at: http://olst.ling.umontreal.ca/~drouinp/termostat_web/index.php

TermoStat also employs Schmid's *Tree Tagger* as lemmatiser and POS (part of speech) tagger thus producing a list where not only is the term's specificity value recorded but also its frequency as lemma, its variants, and its POS tag, as shown in figure 2. The lexical categories identified by *TermoStat* are: nouns, adjectives, adverbs and verbs. It also detects MWTs having nouns and adjectives as phrase heads.

Results

Candidate (grouping variant)	Frequency	Score (Specificity)	Variants	Pattern
section	9694	126.29	section / sections	Common_Noun
v	6828	112.55	v	Common_Noun
case	11465	111.79	case / cases	Common_Noun
para	5973	108.63	para / paras	Common_Noun
article	5686	97.39	article / articles	Common_Noun
court	6387	88.65	court / courts	Common_Noun
appeal	3993	80.3	appeal / appeals	Common_Noun
appellant	3102	78.47	appellant / appellants	Common_Noun
not	22062	75.07	not	Adverb
law	5484	73.55	law / laws	Common_Noun
judgment	2862	71.67	judgment / judgments	Common_Noun
claim	3293	69.8	claim / claims	Common_Noun
right	5795	67.98	right / rights	Common_Noun
apply	3542	65.5	apply / applying	Verb

Figure 2. *Screenshot of output produced by TermoStat*

Based on previous work on lexicon specificity such as Muller's, Lafon's, or Lebart and Salem's (in Drouin, 2003), Drouin claims that the frequency of technical terms in a specialised context differs, in one way or other, from the same value in a general environment and that "focusing on the context surrounding the lexical items that adopt a highly specific behaviour ... can help us identify terms" (Drouin, 2003: 100).

The author uses a corpus comparison approach which provides information on a candidate term's standard normal distribution giving "access to two criteria to quantify the specifity of the items in the set … because the probability values declined rapidly,

we decided to use the test-value since it provides much more granularity in the results"
(Drouin, 2003: 101).

He applies human and automatic validation methods to evaluate the levels of
precision and recall of his software. The author resorts to three specialists who identify
the true terms (TT) from the list generated by *TermoStat* noticing that subjectivity
played a relevant role in this evaluation phase and that it might also be interesting to
study human influence on validation processes. Regarding automatic validation, he
compares the lists of CTs with a telecommunications terminology database. *TermoStat*
reaches 86% precision in the extraction of SWTs. The author insists on the importance
of complementing these methods with others that help identify the meanings of those
words which activate a specialised sense in a specific context.

3.1.4.3. Chung (2003)

On the other hand, Chung's approach to term extraction consists in establishing a
threshold to discriminate terms from non-terms affirming that "to be classified as a
technical term, a type had to occur at least 50 times more often in the technical text than
in the comparison corpus, or only occur in the comparison corpus" (Chung, 2003b: 53).

Chung reaches this conclusion after validating her method by comparison with a
qualitative one, the *rating scale approach,* with the purpose of assessing the degree of
overlap between it and the quantitative technique employed by her. Thus, two experts
are asked to classify the vocabulary in a 5,500 word text from her anatomy corpus, the
sublanguage she analyses in the design and evaluation of her method. They classify the
words into four different categories depending on their level of specialization.

In contrast, the quantitative method employed by Chung consists in calculating
the ratio of occurrence of the word types in the anatomy text given to the experts. The

author normalises the frequencies of the text types in both her anatomy corpus and a general one and calculates the ratio value by dividing the former by the latter. Then, basing her classification on these results and on the frequency figures obtained, she also produces different groups and compares them to the ones by the specialists. The results of the comparison yield 86% average overlap between the author and the experts, especially regarding highly specific words and non-terms.

The author therefore concludes that this ATR method based on statistical data might be reasonably effective, although the last decision to include a word in a given category must be made by the researcher after either consulting the experts or the contexts of occurrence of a given word, since she believes that the most effective approach is the qualitative one. However, it is time-consuming and cannot be applied to large corpora for efficiency reasons.

3.1.4.4. Kit and Liu (2008)

Kit and Liu's (2008) method measures the degree of termhood of SWTs relying on a corpus comparison technique. It aims at studying the different ways words distribute in a specific subject field, namely, in a specialised 8.8 million-word legal corpus called *BLIS* (*Bilingual Laws Information System*) against a general domain using *BNC* as representative of it. Kit and Liu's ATR method focuses exclusively on SWTs, also called *mono-word terms*, basically to avoid "interference from unithood issues" (Kit & Liu, 2008: 206), that is, to prevent such questions as establishing the degree of cohesion between the elements in a grammatical pattern from becoming an obstacle for the calculation of a word's level of specialisation. These authors acknowledge the greater complexity of classifying a mono-word as a term owing to the fact that the structural

information employed to detect the presence of MWTs in a text cannot be applied to SWT automatic mining.

Kit and Liu's method consists in obtaining the rank difference of the vocabulary items in a specialised corpus and a general one "given a domain corpus D (with a vocabulary V_D) to represent a subject field and a balanced corpus B (with a vocabulary V_B) as background, the termhood of a candidate word w is defined as:

$$ʊ(w) = \frac{r_{D\,(w)}}{|V_D|} - \frac{r_{B\,(w)}}{|V_B|}\text{"}$$ (Kit & Liu, 2008: 212).

The application of this formula for the calculation of ʊ-value therefore consists in introducing the rank position (r_D) of a given term in the SC, the specialised one, and normalise it by dividing it by the total number of items in the list, that is to say, in a vocabulary list of 4,500 items, the divisor would be 4,500. After that, the same calculation will be carried out using the normalised datum for the same vocabulary item (r_B) in the RC, the general one. Finally, the normalised value in the RC will be subtracted from the one in the SC obtaining the ʊ-value of the candidate SWT. This result will indicate its level of specialisation thus, the higher it scores, the more specialised it will be considered. Nevertheless, Kit and Liu do not establish a threshold that splits a list into terms and non-terms but rather place words along a termhood continuum "in a way that candidates with a higher termhood value would be pushed to its high end and those with a lower termhood to its low end" (Kit & Liu, 2008: 212).

The method is evaluated using a specialised glossary of legal terms used as gold standard together with a list of TTs extracted from the specialised corpus which were annotated manually by legislators during the drafting of the legal documents in the corpus. The corpus is tokenised (divided into basic text units) removing all the elements that may cause noise. The text units filtered out of the definite list belong to different categories, namely, punctuation marks, numbering items and numerical expressions and

function words. The tokenisation of the corpus results into a list of 8,808,544 tokens which is filtered obtaining a definite one of 13,806 word types.

After comparing their results with those obtained applying Chung's (2003a; 2003b) frequency ratio, they realise that, although the results are similar, it becomes necessary to improve the rank difference calculation to enhance its performance. They propose two alternatives, the second one being slightly more effective. It consists in normalising both the SC and RC ranks using the sum of all the ranks in the respective corpora as the divisor as follows:

$$\sigma_2(w) = \frac{r_{D\,(w)}}{\sum_{w' \in V_D} r_{D(w')}} - \frac{r_{B\,(w)}}{\sum_{w' \in V_D} r_{B\,(w')}}$$

This improved version of the rank difference performs better reaching a precision level of 98.2% on the first 500 CTs (of 12000 evaluated) and 97% on the first 1000, remaining above 90% on the top 20%.

3.1.4.5. *Term Frequency-Inverse Document Frequency (TF-IDF) (Sparck Jones, 1972)*

As opposed to the other four, the *TF-IDF* measure, used in the fields of information retrieval and text mining, does not employ corpus comparison as a means to determine a word's weight. Neither does *RIDF*, its modified version proposed by Church and Gale (1995). *TF-IDF* measures a word's weight by taking into consideration its frequency in a given document and the number of documents it appears in throughout a corpus. A word will display greater weight if it shows high frequency values and appears in fewer documents. As a result, general usage words are ranked lower while more specialised ones tend to appear at higher positions. This measure, or rather more complex versions

of it, is very frequently employed by search engines to rank documents after a user query.

IDF was originally proposed by Sparck Jones (1972) meaning "a giant leap in the field of information retrieval. Coupled with TF ... it found its way into almost every term weighing scheme" (Robertson, 2004:503). Sparck Jones believed that the fact that a word appeared in many documents was not a good indicator of its representativeness within that set of documents. Contrarily, it appeared that those words which occurred in fewer texts might potentially have greater relevance and be more representative of the documents under analysis.

TF-IDF, that is, the result of multiplying *IDF* by a word's frequency in a given document (TF), has evolved throughout time into more sophisticated and complicated measures, as discussed by Robertson (2004). In this study, the classical formula by Sparck Jones will be applied. It is "defined as $-\log_2 df_w/D$, where *D* is the number of documents in the collection and df_w is the document frequency, the number of documents that contain [the word] *w*" (Church & Gale, 1995: 121).

In this study, for the sake of comparison with the lists produced by the other four methods, *TF-IDF* was slightly modified. Instead of resorting to the frequency of a word within a single document in the corpus, which would leave many of the CTs in the other lists out of the rank produced by it (they might not be found in the document selected), after calculating a word's *IDF* value using Sparck Jones' classical formula, it will be multiplied by the normed frequency value[14] of that word in the whole corpus (our adaptation of TF).

---

[14] This value is obtained by dividing a word's raw frequency by the total number of tokens in the corpus and then multiplying it by a scaling factor to obtain more manageable figures due to corpus size (for instance, in a 2.6 million-word corpus, the scaling factor employed is 1,000).

3.1.4.6. Residual Inverse Document Frequency (Church & Gale, 1995)

Finally, Church and Gale (1995) describe *RIDF* using two words to exemplify and justify their method. One of these words is *boycott* (as opposed to *somewhat*), which displays a high *IDF* value "farther from what would be expected by chance (Poisson)[15]" (Church & Gale, 1995: 121). These authors also exemplify how prediction estimates might differ considerably from observed measures such as *IDF*, especially as regards keywords.

They study the behaviour of *boycott* and *somewhat* and come to the conclusion that, while *boycott* (a much better keyword to identify a group of texts on a given topic within a document collection) tends to concentrate in very few documents deviating from Poisson, the observed values for *somewhat* coincide with what would be expected by it.

This is why they propose a new measure based on Sparck Jones' *IDF*, namely, *Residual Inverse Document Frequency* (*RIDF*) to account for such deviations from a chance-based model. It can be calculated by subtracting the predicted *IDF* from the observed measure thus trying to compensate from the deviations observed, especially in the case of good keywords like *boycott. RIDF* is defined as: $RIDF = -\log(df/D) + \log 1 - \exp - tf/D$ (Yamamoto & Church, 1998: 28), where *df* is document frequency (the number of documents a given word appears in), *D* is the total number of documents in the collection and *tf* the frequency of the term in the whole corpus.

---

[15] In the field of term recognition, Poisson's probabilistic model may apply to words which appear independently in a large document collection with a low chance of occurrence.

3.1.5. Method implementation and evaluation

3.1.5.1. Pre-processing and implementation

The major difficulties encountered in the evaluation of these five methods were, on the one hand, establishing a similar process to assess their precision levels and on the other hand, the intrinsic differences existing amongst them. To begin with, Drouin's *TermoStat* (2003) and *Keywords* (2008) are fully automatic and do not require pre-processing, that is, filtering the lists *a priori* to eliminate as much noise as possible. However, Chung's, *TF-IDF*, and specially Kit and Liu's methods need it before producing their lists of CTs.

As part of the pre-processing phase, Chung resorts to Heatley and Nation's (1996) software *Range* to obtain a frequency word type list based on both her anatomy corpus, the SC, and the LOB and Wellington corpora used as RCs. Then, she discards those word types which do not occur in the SC and also eliminates the texts that may contain any vocabulary related to the anatomy field from the RCs in order "to maximise the statistical contrast between the two corpora" (Chung, 2003a: 233).

Kit and Liu's pre-processing procedure consists in tokenising both BLIS and their background corpus, the *BNC*, and filter noise using stop word lists and eliminating alphanumerical elements. After that, they lemmatise the corpus so as to apply their calculations on lemmata, as shown above.

Concerning *UKSCC*, the 193 texts in it were pre-processed with *Wordsmith 5* by Scott (2008) resulting into a list of 27060 word types. Unlike Chung's pre-processing procedure, the legal texts in *LACELL* were not eliminated. Neither was a frequency threshold established prior to the application of Chung's, *TF-IDF*, or Kit and Liu's methods so even hapax legomena and dis legomena were considered with the purpose of maximising the exhaustiveness of the results obtained. *UKSCC* contains 7339 hapax

legomena, that is, vocabulary items occurring only once, which represent 27.12% of the total amount of word types. They include proper names, both English and foreign, such as *Mulliken*, *Kolinsky*, *Jewison* or *Kilmuir;* misspelled words like *spirituall, burmouth, juridicial, tatutory, ntitlement* and also initials and acronyms, i.e. *SIAL, ECHR, BAILII or LJ*.

After obtaining the frequency data of the word types in *UKSCC* with *Wordsmith*, the corpus was filtered using the function word list and baseword list 15 included in Heatley and Nation's (1996) *Range*. They were imported into an excel spreadsheet employing the search function to eliminate the function words and proper names present in *UKSCC*. The percentage of function words detected amongst *UKSCC* word types was low, just 0.99% of the total. As for baseword list 15, it is an ever growing inventory of proper nouns provided with *Range* which led to the removal of 2519 of these elements shrinking the list by 9.4%. Judging by the numbers, the use of proper nouns appears to be a relatively outstanding feature of this legal genre representing almost 10% of the whole corpus (leaving aside those which do not form part of Nation's list and cannot be detected automatically). Undoubtedly, removing them automatically could increase the level of precision achieved regardless of the method employed. However, these proper nouns had to be carefully supervised before removing them since some of them corresponded with initials or acronyms belonging to the specialised vocabulary of the genre like *LJ* (*Lord Judge), QB* (*Queen's Bench),* or *EC* (*European Court*), amongst others.

The filtered list was also used for the calculation of *TF-IDF* and *RIDF* which do not resort to corpus comparison. The frequency lists of word types provided by *Wordsmith 5.0* not only give information about a word's frequency in the corpus (which has been used as TF for this experiment) but also about its distribution throughout it,

that is to say, how many documents within the collection include a given word. Therefore, these were the parameters employed in these cases.

Another pre-processing step taken solely for the implementation of Kit and Liu's method was the lemmatisation of *UKSCC*. It was lemmatised with Schmid's *Tree Tagger*. It resulted into a list of 4,563 lemmata once the function words, proper names and words not found in *BNC* (following their advice in this respect) were carefully filtered. Kilgariff's lemmatised *BNC* list was used as the RC, as stated above. On the other hand, due to the fact that neither *TermoStat* nor *Keywords* require any pre-processing steps, both lists were filtered *a posteriori*. As proof of its efficiency, Drouin's *Termostat* only kept 22 function words (0.94%) and 8 proper names (0.34%) as CTs (out of 2,333), while the keywords list of 3618 items retained 61 function words (1,68%) and 222 proper nouns (6.13%).

Regarding the actual implementation of the five methods, it must be highlighted that both *TermoStat* and *Keywords* are fully automatic tools which can perform all tasks without any human intervention. As for Chung's, *TF-IDF* and Kit and Liu's techniques, excel spreadsheets were used to apply the formulas the authors include in the description of their methods. Once the word type list obtained with *Wordsmith* was imported into a spreadsheet and filtered eliminating function words and proper names, the formulas corresponding to each method were applied to the whole list of word types (the necessary parameters for each calculation were obtained using the *search* function provided by excel). Then, each list was sorted in descending order so that those items displaying the highest values would be ranked at the top of the list. For those methods requiring corpus comparison, *LACELL* was also processed with *Wordsmith* and imported on a different spreadsheet as well as Kilgariff's *BNC* lemmatised lists.

The parameters necessary to apply those methods which are not fully automatic go as follows:

- Chung: Relative frequency in the SC and RC.

- Kit and Liu: rank position in the SC and RC (in descending order) obtained after sorting the candidates according to their normalised frequency in both corpora.

- *TF-IDF*: Normed frequency of candidates in the SC and number of documents they appear in in the whole document collection.

With respect to the parts of speech extracted by each method, the methods designed by Chung or Kit and Liu do not discriminate amongst lexical categories for the identification of terms since they do not resort to POS tagging, neither do *Keywords* or *TF-IDF*. Hence, any part of speech (except of function words which were filtered out) could potentially be regarded as a term depending on the different parameters considered to establish its termhood level. Conversely, Drouin's software applies POS tagging and can be configured to only extract a given part of speech. Nevertheless, it was adjusted to include nouns, adjectives, verbs and adverbs in the process. The validation process shown below is carried out taking into consideration all lexical categories.

3.1.5.2. Defining a gold standard

The results obtained after applying the five ATR methods on *UKSCC* were validated automatically against a legal glossary used as gold standard. Instead of asking specialists to gather a terminology database extracted from the study corpus, four

different legal English glossaries[16] in raw text format were merged and filtered resulting into a list of 10,088 items containing both single and multi-word terms.

Surprisingly and contrary to Nagakawa and Mori's (2002) assumption that 85% of specialised terms are said to be compound, as already justified above, it appears that only 4,157 of 10,088 legal terms (44.78%) are MWTs being distributed as illustrated in figure 3: 3,276 bi-grams (32.47%), 924 tri-grams (9.15%), 239 MWTs formed by four units (2.36%) and 78 (0.77%) with more than four constituents.
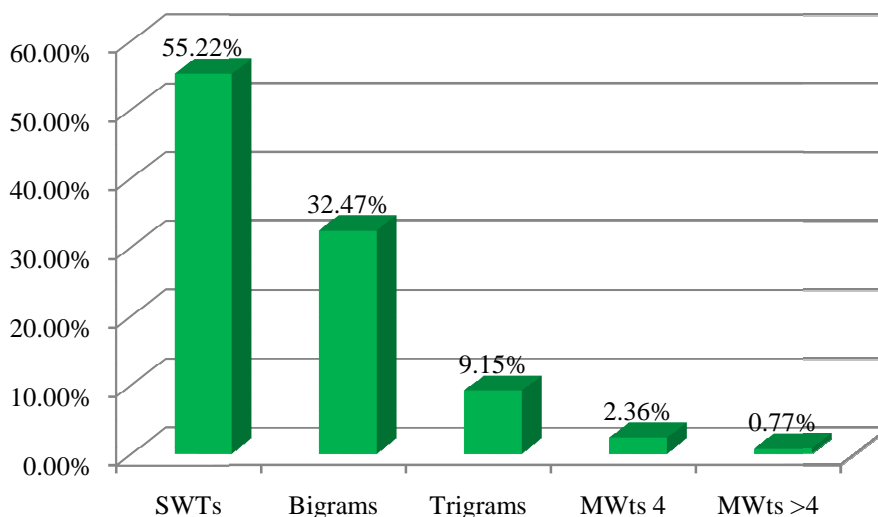
Figure 3. *Lexical structure of terms in glossary*

Once the TTs in each list were identified by comparison with the gold standard, those CTs not qualifying as TTs after applying the methods were analysed manually by the researcher referring to two specialised dictionaries (Alcaraz & Hughes, 1993; Saint Dahl, 1999). This step was taken to contribute to the reduction of silence levels caused

---

[16] Available online at:
http://www.legislation.gov.hk/eng/glossary/homeglos.htm
http://www.judiciary.gov.uk/glossary
http://sixthformlaw.info/03_dictionary/index.htm
http://www.nolo.com/dictionary

by external factors, that is, to guarantee that the glossary, obtained from external sources, would include all the TTs in the corpus. This manual supervision resulted into 10.52% increase of both single and multi-word terms comprised in the glossary list.

3.1.5.3. Results

Defining a similar method of comparison amongst the four approaches under evaluation posed certain difficulties due to the different size of the CT lists produced by each method. While Chung (2003a, 2003b) and Drouin (2003) establish a threshold to discard non-terms, Kit and Liu (2008), *Keywords* (2008), *TF-IDF* and *RIDF* provide a much longer inventory of elements which are ranked according to their level of specialisation. As a result, since Drouin's list included 2,300 items against 4,654 obtained after applying Chung's ratio, 6,675 keywords, and the 27,060 initial word types appearing in *TF-IDF*, *RIDF* and Kit and Liu's lists, only the top 2,000 CTs in each list were selected so that the comparison could be carried out in similar conditions.

These five methods were assessed in terms of precision and recall. Precision can be measured by establishing the proportion of items that are relevant within a given set. This is why it was calculated progressively, as shown in figure 5, where the five curves plot the precision achieved from candidates 1 to 200, 201 to 400, etc. sorted according to the level of specialisation established by each method.

Concerning recall, which points at the amount of TTs identified with respect to the whole list of terms in the corpus (not in a set), it could be calculated for all methods except for Kit and Liu, *TF-IDF* and *RIDF* since neither of them establish a cut-off point to discriminate terms from non-terms. Figure 4 illustrates both average precision and recall.

Nevertheless, Chung's list posed an additional problem which Kit and Liu address when alluding to the items not in the reference corpus. If an item is not in the RC, Chung automatically classifies it as a term and so do Ahmad et *al.* (in Chung, 2003). After examining those elements in BLIS, their study corpus, Kit and Liu (2008: 220) verify that only 20% were TTs and suggest that keeping them "unclassed seems more reasonable when no justifiable solution is available".

Likewise, the number of *UKSCC* items not in *LACELL* was also considerably high, 4,367 single-word CTs were not in the RC and only 280 of them (6.4%) were TTs after comparing them with the gold standard. Thus, it appears that assuming that a word not found in the RC automatically qualifies as a term would not be applicable to our SC, and following Kit and Liu's advice in this respect might be recommendable. As regards the lists produced by the other methods, they do not include these elements either.
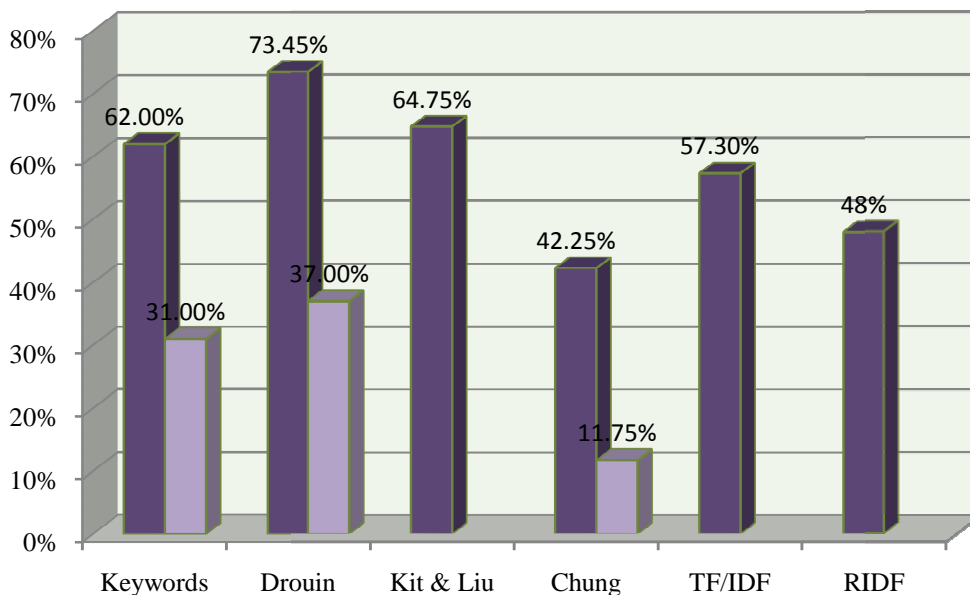
Figure. 4 *Average precision and recall on first 2000 CTs*

As shown in figure 4, the overall precision levels attained by the five methods vary revealing Drouin as the most successful one in identifying terms for this set of 2,000 candidates. It reaches 73.45% being followed by Kit and Liu's which recognises 64.75% of them, *TF-IDF* manages to extract 57.30%, thus ranking fourth. Its modified version, *RIDF*, is 9 points below at 48% while Chung's only identifies 42.25%. As far as *Keywords* is concerned, it ranks third (slightly below Kit and Liu's method) proving to be a considerably effective term extraction tool which detects 62% terms (it reaches 84% precision for the first 200 candidates).

As stated above, calculating recall was not possible for Kit and Liu's method or *TF-IDF* and *RIDF* owing to the fact that the number of CTs coincided with the initial list of word types used to implement the four techniques. Kit and Liu believe that there is no such as thing as a cut-off point and establish a termhood continuum where TTs will be pushed to its high end. The *TF-IDF* and *RIDF* measures do not provide such a cut-off point either.

In general terms, recall figures are not high being Drouin's method the one which excels the other two. It reaches 37% recall followed by *Keywords* at 10 points below. Chung's method is the worst performing one achieving only 11.75%.
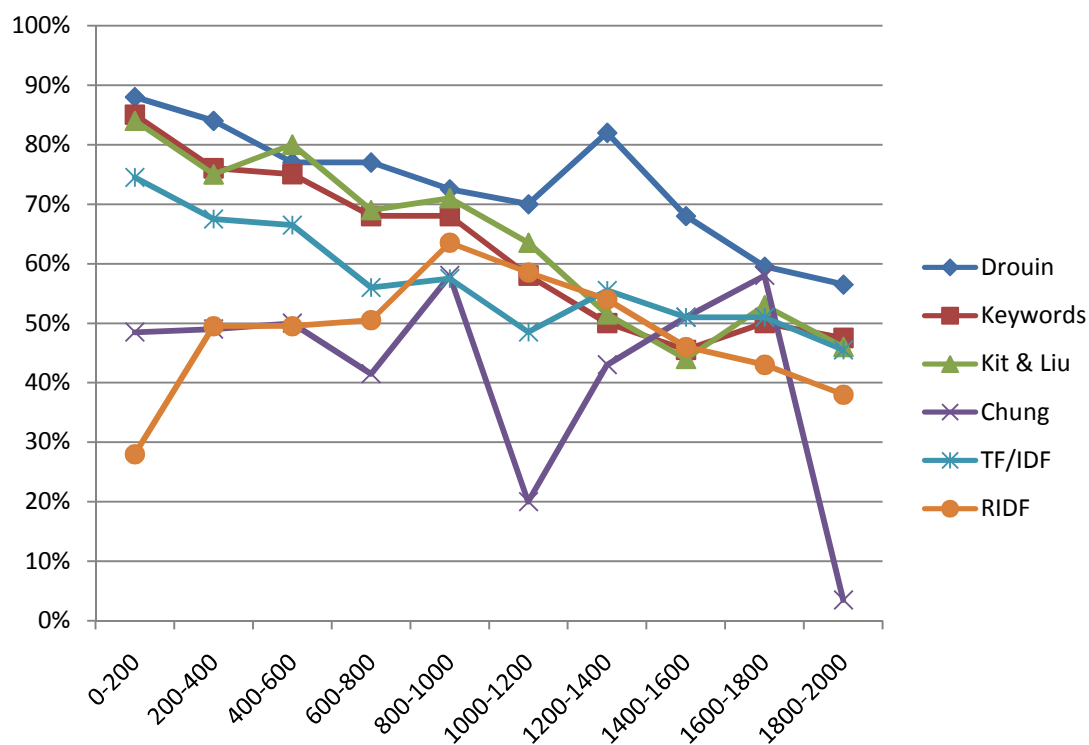
Figure 5. *Cumulative precision for the first 2,000 CTs*

Figure 5 illustrates cumulative precision across methods where the horizontal axis shows the first 2,000 CTs identified in groups of 200 and the vertical one indicates the percentage of precision attained by every method within each group. Drouin's *TermoStat* stands out as the most effective ATR method as it detects 73% terms within the list of 2,000 candidates evaluated. It is closely followed by Kit and Liu's which rises above it only from candidates 400 to 600, where it identifies 80% TTs. The precision levels attained by *Keywords* are reasonably high managing to detect 62% terms (at only 2 points below Kit and Liu). In spite of not resorting to corpus comparison, *TF-IDF* remains considerably close to *Keywords* and Kit and Liu achieving to detect 57% terms for the top 2,000 candidates. Its modified version, *RIDF*, appears to be less efficient as it only extracts 48% TTs from the corpus being the least effective of all. Chung's

method reaches 50% precision on average within this range although it stands below *RIDF* from candidates 400 to 1,400.

TermoStat, *Keywords,* Kit and Liu and *TF-IDF* follow a similar trend decreasing their effectiveness smoothly from candidates 1 to 900. Within this range, Drouin achieves 82% precision, Kit and Liu 77%, *Keywords* 76%, and *TF-IDF* 66% (finding their highest points at 88%, 84%, 84.5%, and 74.5% respectively). On the other hand, Chung remains steady below 50%. Conversely, although it shows really poor results from candidates 0 to 200 (where noise levels reach 76%), *RIDF* increases its effectiveness reaching its peak at 64% within this group.

From candidates 900 to 1,700 there are greater differences. While *TermoStat* remains ahead reaching a peak of 82% within candidates 1100 to 1300 and then falling down to 60%, Kit and Liu, *Keywords, TF-IDF* and *RIDF* continue to descend progressively (more sharply in the case of Kit and Liu) to 53%, 50%, 54% and 46%. On the contrary, Chung improves considerably rising to 58%.

Finally, both Kit and Liu and *TermoStat* fall down to 46% and 56.5% while *Keywords* and *TF-IDF* remain constant at 47.5% and 46% from candidates 1700 to the end of the graph. *RIDF* falls considerably to 38%. The case of Chung's method is particularly outstanding as it falls sharply from 58% precision to 3%. It must be emphasised that the 2,000 candidates considered for evaluation do not correspond with what Chung would regard as terms proper. The cut- off pointed suggested by the author would only apply to the first 287. All the same, the average level of precision within this set does not even reach 50%.

On the whole, having compared and assessed the five methods above, there are several generalisations that could be made with respect to their effectiveness in

extracting terms in a legal English corpus. To begin with, it appears that resorting to corpus comparison yields better results. As a matter of fact, *TermoStat* and *Kit Liu's* methods, the best performing ones, employ this technique to establish a word's termhood level. As regards precision within the list of 2,000 candidates evaluated, both of them stand at 16.15 and 7.5 points above *TF-IDF* and at 25.45 and 16.75 above *RIDF* respectively, which focus exclusively on the specialised corpus to extract CTs.

Another factor that may have influenced their greater rate of success is the fact that, unlike the rest of the methods, both require lemmatisation to be implemented thus indicating that applying calculations on lemmata, not on word types, might be more effective to recognise terms automatically.

Concerning the gold standard employed for evaluation, the fact that it was compiled using external sources does not seem to have affected the results significantly. While Drouin employs a database external to the corpus to assess their method, Kit and Liu resort to a glossary obtained from the texts themselves. However, both methods perform quite efficiently for this study being *TermoStat* the most effective one. Even so, there is not enough evidence to relate Kit and Liu's slightly lower rate of success with the fact that the gold standard was not obtained from the legal corpus itself.

To conclude, it must be highlighted that the low precision levels achieved by Chung's method might point at its domain dependence. As put forward by Lemay et *al*. (2005: 233), "lexical units in medical texts bear certain surface-level features (i.e. morphemes or entire words borrowed from Latin and Greek) that, we believe, make them less difficult to identify automatically". Unlike Chung, who resorts to human validation, the use of a gold standard to automatically validate the results in this experiment could have also contributed to the lack of precision of this method.

Table 2 illustrates the first 25 CTs detected by each method ranked in descending order from higher to lower termhood levels according to the different measures proposed by each author.

Table 2

First 25 CTs ranked by every ATR method

DROUIN		KEYWORDS		KIT & LIU		CHUNG		TF-IDF	
SECTION	126.29	COURT	27965.27	COURT	0.3114	CRAIGHEAD	2198.45	LAND	0.998
V (VERSUS)	112.55	SECTION	24182.76	JUDGE	0.3110	APPELLANTS	2012.58	ARTICLE	0.965
CASE	111.79	PARA	22007.62	CASE	0.3105	CIV	1846.69	CONTRACT	0.926
PARA (PARAGRAPH)	108.63	LORD	21963.51	SENTENCE	0.3100	APPELLANT'S	1577.55	JEWISH	0.898
ARTICLE	97.39	V	19464.25	CONTRACT	0.3095	PARAS (PARAGRAPHS)	1444.69	EXTRADITION	0.866
COURT	88.65	APPEAL	18886.16	APPEAL	0.3091	COBBE	1079.23	POSSESSION	0.861
APPEAL	80.3	ARTICLE	18044.94	TERM	0.3086	ESTOPPEL	975.31	CHILD	0.845
APPELLANT	78.47	ACT	17322.12	JUDGMENT	0.3081	LESSEE	639.54	TENANT	0.804
LAW	73.55	CASE	16541.39	MAKE	0.3076	PPC	607.57	COMPANY	0.783
JUDGMENT	71.67	LAW	10566.68	ISSUE	0.3072	RESPONDENT'S	591.58	CONVENTION	0.775
CLAIM	69.8	JUDGMENT	8741.90	ORDER	0.3067	APPELLANT	582.05	ASYLUM	0.724
RIGHT	67.98	CONVENTION	7648.50	OFFENCE	0.3062	REALISABLE	567.59	DATA	0.721
APPLY	65.5	RIGHTS	7304.34	APPELLANT	0.3057	LAWFULNESS	563.60	DIRECTIVE	0.702
ORDER	64.39	WHETHER	7262.35	COSTS	0.3053	TORTIOUS	559.60	EQUIPMENT	0.701
DECISION	63.53	DECISION	7056.68	MONTH	0.3048	SENESCHAL	535.62	IMMIGRATION	0.656
PERSON	62.83	APPELLANT	6947.53	TAKE	0.3043	PARA (PARAGRAPH)	530.02	DISCRIMINATION	0.647
PROCEEDING	61.7	PROCEEDINGS	6927.94	TRIAL	0.3038	CARNWATH	519.63	SUICIDE	0.645
RELEVANT	59.02	LJ	6707.16	SAY	0.3034	DISAPPLICATION	495.65	RENT	0.645
PURPOSE	58.45	JURISDICTION	5968.92	EVIDENCE	0.3029	STEYN	491.65	ACCOMMODATION	0.627
DEFENDANT	57.72	ORDER	5762.57	SUSPENDED	0.3024	FORESEEABILITY	439.69	PLANNING	0.614
PROVISION	57.55	RELEVANT	5427.42	DEFENDANTS	0.3019	INTERVENERS	439.69	CRIMINAL	0.614
PRINCIPLE	55.77	AC	5071.04	FACT	0.3015	ABBOTSBURY	401.71	COMMISSIONERS	0.608
APPLICATION	55.5	PARAS (PARAGRAPHS)	5051.25	CONCLUSION	0.3010	SUBSECTION	384.79	CLAUSE	0.583
JURISDICTION	55.5	APPLICATION	4801.27	GIVE	0.3005	NUPTIAL	373.73	PROPERTY	0.580
PARAGRAPH	54.69	KINGDOM	2796.27	REASON	0.3000	INVERESK	371.73	LEASE	0.576

RIDF			DALLAH	6.4986
NUPTIAL	6.9039		TULLIS	6.4183
BIOT	6.6650		CHAGOSSIANS	6.3186
STOJEVIC	6.6161		IMGS	6.2638

ASCO	6.2409		AUDITORS	5.8462
OCR	6.2174		GHALANOS	5.8407
DONATIONS	6.2054		SENESCHAL	5.8229
INVERESK	6.2054		ALLDECH	5.8047
DISCREETLY	6.0619		OTHMAN	5.7862
SAINSBURY'S	6.0332		PETER'S	5.7862
AUDITOR	5.9255		TAXOL	5.7862
HSMP	5.9091		STEART	5.7482
ECRC	5.8583		SEWER	5.7292

3.1.5.4 Processing of BLaRC: identification of SWTs

Having validated six different SWT recognition methods and the output vocabulary lists produced by each of them, the most efficient one, Drouin's (2003) *TermoStat*, was selected to process *BLaRC*, the 8.85 million-word legal corpus. First, it was implemented to extract the list of SWTs from the whole corpus according to the parameters described above for this method.

As far as precision is concerned, *Termostat* establishes a specificity threshold at 3.09 which defines a cut off point leading into the recognition of 2200 CTs. 1564 of these terms were validated as TTs thus reaching 71% precision on average. 636 out 1564 candidates were false positives, that is, 29%, while the percentage of true negatives was considerably high: 84% (25049 out of 29489). Lastly, only 4400 TTs remained undetected leading to 16% silence, as illustrated in figure 6.
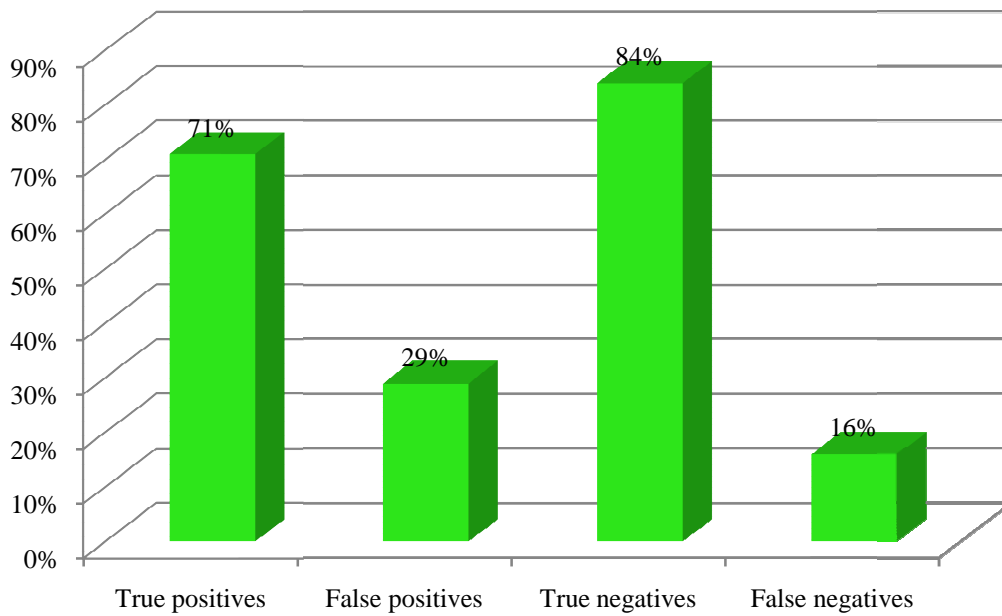
Figure 6. *True and false positives, true and false negatives identified in BLaRC*

using Termostat

Having analysed precision after grouping the CTs in sets of 200 items, the levels

achieved for each group go as follows:
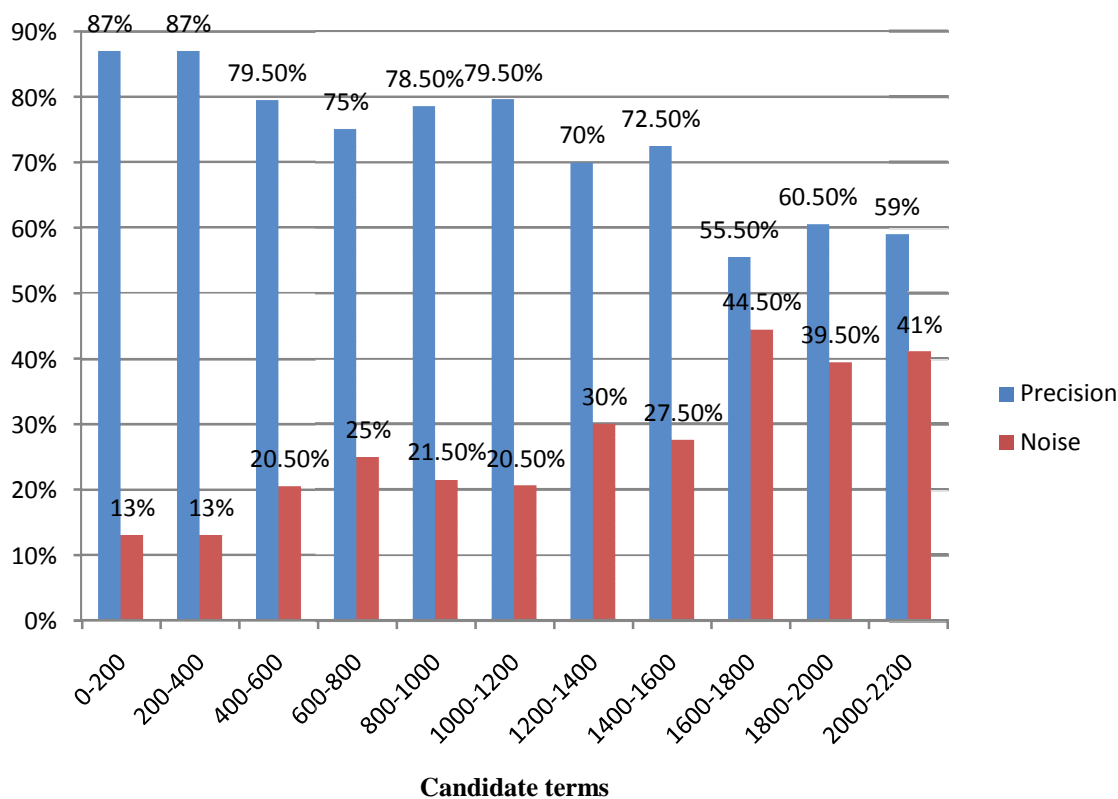
Figure 7. *Cumulative precision obtained by TermoStat after applying it on BLaRC*

As shown in the graph above, precision reaches almost 90% for the top 400 CTs while it remains steady at 8 points below within the next group of 800. The method identifies 70% TTs from candidates 1200 to 1600 decreasing sharply its efficiency to 56% from CTs 1600 to 1800. Finally, it improves precision going up to 60% from candidates 1800 to 2200. Noise levels remain inversely proportional to precision from 10% for the top 400 candidates identified to 40% for the last 600 items in the list.

Recall was also measured for this method since it clearly provides a cut off point to discriminate terms from non-terms. On a whole, *Termostat* reaches 26% recall, that is, the total amount of TTs identified in the corpus with respect to the whole list of terms in the corpus. It was measured by dividing the number of TTs extracted by the

method by the total amount of terms in the corpus (including false negatives), then the result was multiplied by 100 to be able to express it as a percentage.

Nevertheless, if Drouin's specificity threshold was lowered from 3.09 to 1.34 (including the next 2000 items in the list), recall would increase to 38.9% (managing to identify 2339 TTs). However, average precision diminishes in inverse proportion to recall dropping from 71% to 55%. If the threshold continued to be reduced (to 0.66), recall would augment considerably to 63% affecting precision, which would decrease to 45.12%.

Finally, table 3 presents the list of SWTs extracted from *BLaRC* by *Termostat* once it was validated using the gold standard for comparison. It includes those terms whose level of specialisation was above the threshold established by Drouin, 3.09. 2522 of the single-word candidate terms retrieved by the method coincided with part of the list produced after implementing Nazar and Cabré's (2012) *Terminus 2.0* on our corpus. Thus, they were removed from the list below being only included in the term inventory obtained with *Termimus,* shown in table 4.The resulting list contains 541 SWTs.

Table 3

List of SWTs in BLaRC identified by Termostat

TERM	SPECIFICITY		
CASE	89.24	PURPOSE	43.06
V (VERSUS)	67.43	PROVISION	42.15
DECISION	63.86	RIGHT	40.6
PARAGRAPH	61.09	CONCLUSION	39.12
PARA (PARAGRAPH)	52.55	GROUND	37.02
RELEVANT	50.55	STATEMENT	37
PROCEEDING	48.54	APPLICANT	36.25
SUBMISSION	44.65	NOTICE	35.51
FACT	43.58	PARTY	35.16
LAW	43.1	OFFENCE	33.99
		PRINCIPLE	33.94

TERM	33.75	LORDSHIP	18
AGREEMENT	32.55	ALLEGED	17.89
STATUTORY	32.42	LEGISLATION	17.81
CONTRACT	31.34	FAIR	17.76
SENTENCE	30.96	EMPLOYEE	17.72
LIABILITY	30.92	CONTRACTUAL	17.69
TRIAL	29.81	DISCRIMINATION	17.66
OPINION	29.61	LANDLORD	17.5
CONSIDERATION	29.54	ACTION	17.48
PROPERTY	29.41	AMENDMENT	17.44
REVIEW	29.19	ASYLUM	17.4
FINDING	28.84	DECEASED	17.39
DUTY	28.61	INVESTIGATION	17.25
REGARD	28.17	FACTUAL	16.96
REQUIREMENT	28.1	PROCEDURAL	16.92
PARTICULAR	26.69	SUBSTANTIVE	16.65
ACCORDANCE	26.59	DISCIPLINARY	16.63
CLAUSE	26.44	LITIGATION	16.53
DEFENDER	26.25	HARM	16.41
PERMISSION	26.09	APPLICABLE	16.26
EMPLOYER	25.47	EXTRADITION	16.26
POSSESSION	25.37	LICENCE	16.19
REGISTRATION	24.89	NEGLIGENCE	16.09
DISCLOSURE	24.41	COMPENSATION	16.07
ALLEGATION	24.38	PROOF	16.07
JURY	23.99	FRAUD	15.34
SUBSECTION	23.24	DIRECTION	15.25
ASSESSMENT	22.29	INTERPRETATION	15.14
UNLAWFUL	22.07	CROSS-EXAMINATION	15.03
LEAVE	21.59	PANEL	15
PURSUANT	21.51	RELIANCE	14.95
PROSECUTION	21.14	CONFISCATION	14.55
REASONING	20.73	ABUSIVE	14.52
PARAS	20.57	CONSISTENT	14.51
PLAINTIFF	20.49	LEGITIMATE	14.39
DETERMINATION	19.71	SUBSTANTIAL	13.94
DETENTION	19.44	CERTIFICATE	13.93
STATUTE	19.12	DEFENCE	13.87
ASSET	18.85	THEREAFTER	13.66
DISPUTE	18.53	CREDITOR	13.58
MATERIAL	18.41	LIMITATION	13.49
UNFAIR	18.02	DECLARATION	13.29

PRISONER	13.15		PROPRIETARY	10.14
BIND	13.07		SENTENCING	10.12
TORT	13.06		NEGLIGENT	10.1
CORRESPONDENCE	12.98		DISPUTE	10.09
UNDERTAKING	12.98		INJURY	10.06
REPAYMENT	12.61		LEGISLATIVE	9.98
EXCEPTIONAL	12.5		PUBLIC	9.89
PROPORTIONALITY	12.49		ANCILLARY	9.88
SUBSEQUENT	12.48		ADVOCATE	9.79
SUBSEQUENTLY	12.42		COMPLAINER	9.78
PLEA	12.29		OBJECTIVE	9.7
RESIDENCE	12.03		PETITIONER	9.66
REPRESENTATION	11.83		CORONER	9.61
FRAUDULENT	11.77		REINSURANCE	9.61
REMUNERATION	11.71		COMMENCEMENT	9.57
RECEIPT	11.55		LIQUIDATOR	9.56
LIBERTY	11.3		AFFIDAVIT	9.47
NATIONALITY	11.29		DECISION-MAKER	9.39
OFFENDER	11.28		INSTRUCTION	9.32
INSURER	11.28		ENFORCEMENT	9.28
WARRANT	11.23		FORFEITURE	9.28
NOTIFICATION	11.13		AUTHORISATION	9.21
EQUITABLE	10.89		ER	8.99
DOCUMENTATION	10.78		DISCRETIONARY	8.97
COMPLIANCE	10.75		ADJUDICATION	8.94
RELIEF	10.71		DEDUCTION	8.92
COUNTERCLAIM	10.69		ADMISSIBILITY	8.91
RESOLVE	10.66		INDEMNITY	8.87
RECONSIDERATION	10.64		REGIME	8.79
CONSPIRACY	10.6		ADMISSION	8.71
IMPLIED	10.58		CROSS-EXAMINE	8.69
MATRIMONIAL	10.56		REQUISITE	8.67
GUILTY	10.52		INSURANCE	8.64
TRADER	10.44		EVIDENTIAL	8.63
DEBTOR	10.42		RESPONDENT'S	8.63
HEARSAY	10.35		TRIBUNAL'S	8.61
DILIGENCE	10.32		COMPETENT	8.58
EXCEPTION	10.31		UNREPORTED	8.58
JJ	10.26		VALIDITY	8.57
TAX	10.23		TAXABLE	8.53
CUSTODY	10.22		REVIEW	8.51
DEPORTATION	10.22		REG.	8.5

INSOLVENCY	8.49	MANDATORY	7.3
BENEFICIARY	8.42	UNJUST	7.29
RESTRAINT	8.4	ACCUSE	7.27
PROVISIONAL	8.37	ASSIGNMENT	7.26
ASSURANCE	8.33	ACQUISITION	7.23
RELEASE	8.31	DWELLING	7.2
BINDING	8.29	SIC	7.2
COMPLETION	8.22	COPYRIGHT	7.12
JUROR	8.2	LESSEE	7.09
EXEMPTION	8.18	CONTRACTING	7.08
SENTENCE	8.16	NON-RESIDENT	7.07
OWNER	8.15	DETRIMENTAL	7.04
APPELLANT'S	8.13	GENUINE	7.04
MATERIALLY	8.11	PRIVILEGE	7
EXCLUSION	8.08	SUITABILITY	6.98
RECOVERY	8.05	MITIGATION	6.97
NON-DISCLOSURE	8.04	TAXPAYER	6.96
SUB-PARAGRAPH	7.99	DEPUTE	6.95
SERIOUS	7.98	INVALID	6.94
CONSEQUENTIAL	7.97	SELF-EMPLOYED	6.87
ACCOUNTING	7.95	SECURITY	6.86
REGISTERED	7.92	REVENUE	6.84
ADJUSTMENT	7.89	ALLOWANCE	6.81
FIDUCIARY	7.84	COMPULSORY	6.81
RESIDENT	7.8	ARBITRAL	6.78
LIQUIDATION	7.79	CHARGEABLE	6.78
PRESCRIBED	7.77	TERRORIST	6.68
SCHEME	7.75	INCAPACITY	6.66
GUARANTOR	7.73	SAFEGUARD	6.66
PREMIUM	7.68	REPORT	6.65
SCRUTINY	7.61	DEPRIVATION	6.62
SUBSECTIONS	7.6	NOTIONAL	6.62
PROPRIETOR	7.59	AGGREGATE	6.6
DISPUTED	7.58	TRESPASS	6.58
QUANTUM	7.44	FUNDAMENTAL	6.57
SUB-SECTION	7.39	QUALIFICATION	6.52
UNPAID	7.36	FALSE	6.51
VALUER	7.35	SUBJECT-MATTER	6.51
CONDITIONAL	7.33	HARASSMENT	6.48
TRANSITIONAL	7.32	SIAC	6.48
CONFIDENTIAL	7.31	SAFEGUARD	6.45
UNAUTHORISED	7.31	PENALTY	6.44

FSA	6.39		HERITABLE	5.71
SCHEDULE	6.39		ILLEGALITY	5.7
NIL	6.38		IMPLIEDLY	5.7
WAIVER	6.38		INTERDICT	5.7
RETROSPECTIVE	6.36		MEASURE	5.7
VICTIMISATION	6.36		REGISTRANT	5.7
CONTINGENT	6.35		SUMMARILY	5.7
SURVEILLANCE	6.33		CLEARANCE	5.64
CONTRIBUTORY	6.3		BEARING	5.63
MAGISTRATE	6.26		CROSS-APPEAL	5.63
PRECISE	6.26		DEFECTIVE	5.63
DISPOSAL	6.22		DECEASE	5.61
SUSPENSION	6.21		REGARD	5.61
PROCESS	6.17		SUBSTITUTION	5.6
IDENTIFICATION	6.16		EXCISE	5.57
VEST	6.11		NON-COMPLIANCE	5.57
JOINT	6.09		RECOUPMENT	5.57
QUALIFIED	6.07		ISSUE	5.56
REHEARING	6.07		MATTER	5.55
SHAREHOLDING	6.07		JUDGE'S	5.53
SUBSIST	6.07		SUPERVISION	5.53
DEFAMATION	6.04		MISCARRIAGE	5.51
SANCTION	6.04		RAPE	5.51
PROCESSION	6.03		INTERVENER	5.46
PRICE-FIXING	6.01		COMPLIANT	5.44
OVERRIDING	5.97		RESIDUARY	5.42
REMARK	5.96		ALIEN	5.41
RESOLUTION	5.96		HOLDER	5.41
NEGLIGENTLY	5.95		HOMELESSNESS	5.39
VERDICT	5.95		BAR	5.38
CASE-LAW	5.9		CONFIDENTIALITY	5.38
CAPACITY	5.86		COLLATERAL	5.34
EXECUTION	5.86		RULING	5.31
COMPENSATORY	5.84		HOUSING	5.27
CONTRA-TRADERS	5.82		LIEU	5.27
REVOCATION	5.82		ADVERSELY	5.26
RE-HEARING	5.79		SURVEYOR	5.25
DEROGATION	5.77		WARRANTY	5.25
ANTE-NUPTIAL	5.76		VICARIOUS	5.24
DECREE	5.76		PRE-HEARING	5.23
EXECUTOR	5.74		PRE-SENTENCE	5.23
BANKRUPTCY	5.71		COHABITATION	5.2

CHARTERPARTY	5.19	VERIFICATION	4.57
PREMISES	5.18	RE-OFFENDING	4.56
JOINDER	5.16	INTENTIONALLY	4.54
COMPULSORILY	5.15	LIBEL	4.54
DISSENTING	5.13	LICENSING	4.54
CAUTION	5.11	EVICTION	4.53
ACQUITTAL	5.09	ALIVE	4.52
LEASEHOLD	5.06	RETENTION	4.52
RECTIFICATION	5.02	APPORTIONMENT	4.46
ACCESSION	5.01	DECEIT	4.46
TRANSFEROR	4.99	DECEPTION	4.46
CONTRACTUALLY	4.97	MORTGAGE	4.43
LOAN	4.97	INSURED	4.42
COMMENT	4.94	CONTRACTING	4.39
IMGS	4.94	SURCHARGE	4.39
OFFICER	4.94	CREDIT	4.38
CO-DEFENDANT	4.93	FACT-FINDING	4.37
REGULATORY	4.92	PROFIT	4.37
ESTATE	4.89	SUPPORT	4.37
DECLARATORY	4.88	FORENSIC	4.36
NON-EXCLUSIVE	4.87	FILE	4.35
GROSS	4.86	INSTRUCTED	4.35
BARONESS	4.85	MANSLAUGHTER	4.35
DEBT	4.85	MEMORANDUM	4.35
VEHICLE	4.85	SALE	4.35
PREAMBLE	4.82	SEVERALLY	4.35
SUMMING-UP	4.82	INVESTIGATORY	4.33
EQUITY	4.8	ACCEPTANCE	4.32
PENAL	4.79	IMMUNITY	4.32
MODIFICATION	4.78	WARRANT	4.31
SUMMONS	4.76	PROCEDURALLY	4.3
NON-PAYMENT	4.73	FRAUDULENTLY	4.27
SIC	4.73	RESTITUTION	4.25
BODILY	4.72	WRIT	4.24
DURATION	4.72	ASSURED	4.22
GOODWILL	4.71	AUDIT	4.21
LICENSEE	4.7	REQUISITION	4.2
LICENSE	4.65	DISORDER	4.15
DISAPPLICATION	4.62	LIQUIDATE	4.14
PETITION	4.61	WAYLEAVE	4.14
IPT	4.6	LOCALITY	4.13
POST-NUPTIAL	4.6	FORTHWITH	4.12

PLANNING	4.11	GOVERNANCE	3.65
TREATMENT	4.11	RECITAL	3.64
MAXIM	4.1	EXCISE	3.63
CONVEYANCING	4.09	RESTITUTIONARY	3.62
EXTRA-TERRITORIAL	4.09	CANVASS	3.6
ELIGIBILITY	4.08	CITIZENSHIP	3.59
VIZ.	4.08	CONVENE	3.59
INSTITUTE	4.07	ARP	3.57
PROSCRIBE	4.07	FAULT	3.56
PRE-TRIAL	4.05	NON-DISCRIMINATORY	3.56
COMPETENCE	4.03	COUNTERVAIL	3.55
DEMOTE	4.03	AGGREGATE	3.53
MISUSE	4.01	CONTESTED	3.53
CONTRACT	3.97	NEAR-SUICIDE	3.51
ACCOUNTANT	3.92	PRE-CONDITION	3.51
BARRISTER-AT-LAW	3.91	STATUTE-BARRED	3.51
GUARDIAN	3.91	TRANSPARENCY	3.5
PRIVACY	3.91	TRANSFEREE	3.48
REPURCHASE	3.91	FIRST-NAMED	3.46
REPEAL	3.9	WARRANDICE	3.46
INDECENT	3.88	DOMICILE	3.45
FOSTER	3.87	BAIL	3.42
DEFER	3.86	RE-TRIAL	3.41
PRE-ACTION	3.86	THEREUNDER	3.41
ASYLUM-SEEKER	3.85	HANSARD	3.4
REVIEW	3.84	ACKNOWLEDGEMENT	3.38
UNSECURED	3.82	INTENTIONAL	3.37
CAUSAL	3.81	IN-COUNTRY	3.36
CO-ACCUSED	3.81	NON-PARTICIPATING	3.36
DELICT	3.81	PLEA-IN-LAW	3.36
PROBATION	3.81	TRANSNATIONAL	3.34
CONSULAR	3.8	OVERDRAFT	3.32
CORROBORATE	3.79	PARAMOUNT	3.32
TRADING	3.79	DISCUSSION	3.3
EXCISE	3.77	IA	3.3
BENEFICIALLY	3.75	LESSOR	3.3
FAIR-MINDED	3.75	SUB-CONTRACTORS	3.3
ENFRANCHISEMENT	3.72	SUBSTANTIVELY	3.3
OVERSEAS	3.71	DISTRESS	3.29
EXPEDITIOUSLY	3.67	MISUNDERSTANDING	3.28
TRAVAUX	3.67	ARBITRARY	3.25
RESIDUAL	3.66	LENDER	3.25

BAILMENT	3.24		INSURED	3.17
EXPLANATORY	3.24		AMENDED	3.13
PURSUANT	3.24		FIRST-TIER	3.13
FACTUALLY	3.23		PROSPECTIVELY	3.12
INSPECTION	3.21		RULE-MAKING	3.12
ABSOLVITOR	3.19		SUPPLEMENTAL	3.12
CONSULTANT	3.19		NON-COMMERCIAL	3.07
ALLEGEDLY	3.17			

3.2. SWT AND MWT RECOGNITION METHODS

3.2.1. Introduction

Once Patrick Drouin's (2003) *TermoStat* has been selected as the most efficient SWT recognition method and implemented on *BLaRC*, a validation process similar to the one applied to SWT recognition methods will be carried out to single out the most effective method capable of recognising MWTs automatically.

A brief description of the latter type of methods is provided below together with an explanation of their implementation on the pilot 2.6 million-word corpus, *UKSCC*. The results obtained after their implementation will be studied and the best performing one, Nazar and Cabré's (2012) *Terminus,* will, in turn, be applied to the 8.85 million-word corpus, *BLaRC*. Finally, the resulting list of MWTs obtained will also be presented.

Most of these methods can recognise both SWTs and MWTs so their validation will include precision levels for both types of units (in case they were capable of extracting both as it happens with *Terminus* (Nazar & Cabré, 2012))*, Termextractor* (Sclano & Velardi, 2007) and *Textract* (Park et *al.*, 2002).

3.2.2. Method description

3.2.2.1. Terminus 2.0 (Nazar & Cabré, 2012)

Nazar and Cabré propose an ATR method, freely available online[10], whereby term extraction becomes a fast and easy task. *Terminus 2.0* offers different possibilities for the researcher working on specialised terminology. As indicated on the website guide, it has varied functions such as textual corpus search, compilation and analysis; term extraction; glossary and project management; database creation and maintenance and dictionary edition.

Their ATR method is based on the assumption that the system can learn how to recognise terms based on the language samples provided by the user. The expert does not need to formulate rules to help the system work but rather let it learn from the real samples provided of both specialised terms and general language using the latter for comparison.

The program "develops a statistical model with an abstraction of the main characteristics of both samples" (Nazar & Cabré, 2012: 210). As it is open to any user who can upload glossaries and corpora to help the system learn to identify terms in different domains, the more users employ it, the greater its capability will become to identify terminological units. As stated by the authors, the greatest innovation of this method is its collaborative character since it "allows a community of terminologists to share knowledge acquired by the program in each training phase" (Nazar & Cabré, 2012: 212).

The method applied by the system is structured into three distinct phases: syntactic, lexical and morphological. To begin with, using Schmid's *Tree Tagger* (1995), the texts are POS tagged and a syntactic model is developed based on the

---

[10]Available at: http://terminus.upf.edu

frequency of distribution of the syntactic patterns identified. After doing so, the frequency of the lexical units displaying those patterns is measured. Finally, it extracts initial and final character *n*-grams. The termhood score is obtained by assigning a higher value to those units which have a "significant frequency in the LSP training material with respect to the general language corpus" (Nazar & Cabré, 2012: 212). This process is followed for all levels of training.

The authors act as judges to validate their method by confirming the candidates extracted as TTs and discarding those which do not qualify as such. The corpus employed as the training set is a 300,000 word collection of papers on corpus linguistics (CL) published in 2010. The test corpus is also a collection of papers on the same topic of similar size (340,000 words). Both sets of texts were taken from the scientific journal *Computational Linguistics.* The reference corpus consists in a 2 million-word collection of press articles from the Leipzig Corpora Collection. In the evaluation process the algorithm is trained also using *n*-gram frequency lists and word association measures.

As part of this training, the authors validate 800 terminological units and train the algorithm using this list of terms (both SWTs and MWTs). Once the training phase is accomplished, the study corpus is processed employing the information derived from the training. For the validation of the results obtained after processing the study corpus of 340,000 words, the authors resort to three different classical measures, namely, chi-square test, mutual information and frequency (the most frequent 1500 bigrams are extracted). They also employ a stop word list to filter the results.

As a result, the precision levels achieved are considerably better than those attained by the three methods used for comparison. *Terminus* reaches 85% precision for the top 200 candidates and 75% for the top 400.

3.2.2.2. C-value (Frantzi et. al., 1999)

This ATR method does not resort to corpus comparison but rather stands as a domain-independent one only based on a specialised corpus. It is a hybrid method which employs both linguistic and statistical data to produce a list of CTs ranked according to their termhood score. A term's c-value can be calculated with respect to its frequency and the frequency of its sub-terms:

$$CValue(a) = \log_2 |a| \cdot \left(f(a) - \frac{1}{P(T_{a)}} \sum_{b \in T_a} f(b) \right)$$

Where, $f(a)$ is the frequency of term (a) with $|a|$ words, T_a is the set of CTs recognised by the method that contain (a) and $P(T_{a)}$ is the total number of longer CTs that contain (a).

The linguistic part of the method is articulated into different steps which go as follows:

1- The corpus is POS tagged.

2- A linguistic filter is applied so as to discard certain patterns and keep a balance between precision and recall (the use of an open filter could favour recall at the expense of precision). Only those strings containing nouns premodified by other nouns, adjectives or combinations of both are kept.

3- A stop list is employed which comprises both function words and high frequency ones from a sample corpus not expected to be terms.

As part of the statistical parameters utilised to select the CTs, the authors take into consideration the frequency of occurrence of the pattern, also the frequency of the pattern as part of other longer structures, the amount of these longer structures and the number of constituents of the pattern.

Frantzi et *al*. introduce the concept of *nested terms* as key within the statistical part of their method. With the purpose of trying to discard those patterns which are not TTs, they decide to select only those which contain strings which also appear by themselves in the corpus displaying relatively high frequency. A frequency threshold of >3 is applied to avoid producing a too long list that might become a hindrance for the experts evaluating the output.

For the assessment of their method, the authors highlight the fact that there is no agreement amongst experts and that such subjectivity necessarily leads to the introduction of the concept of 'relative' precision and recall. Instead of asking an expert to extract all the terms in a corpus, which is time-consuming and hard to attain, recall figures are obtained "with respect to frequency of occurrence, which we use as the baseline method" (Frantzi et *al.,* 1999: 8).

The authors also assess precision at three stages: first, evaluating those candidates which have appeared as nested; second, evaluating only those appearing as nested, and third, evaluating all the CTs. As a result, the authors realise that, in general, the use of a more open linguistic filter does not affect precision significantly. Moreover, using other statistical data "apart from the pure frequency of occurrence of CTs, improves the precision of the extracted nested multi-word terms, with a slight only loss on recall" (Frantzi et *al.,* 13).

3.2.2.3. TermExtractor (Sclano & Velardi, 2007)

Sclano and Velardi's method introduces an evaluation process different from other ATR methods. The results obtained by *TermExtractor[11],* the free online tool developed by the authors, are assessed "by web communities and individual users on different domains"

---

[11] Available at http://lcl.uniroma1.it/termextractor

(Sclano & Velardi, 2007: 6). The online software interface allows the creation of a team of judges who will validate the results obtained once a given corpus has been processed and a list of CTs produced. The average precision attained having consulted both private and public institutions (such as Stockholm University, the University of Ottawa or the Institute of Systems Analysis and Computer Science in Rome, amongst many others), as well as private users, was 80%, reaching a peak of 99.4% for a group of texts (7680) belonging to the field of anatomy and medicine.

TermExtractor manages to identify terms based on two distinct phases. The first one, linguistic, consisting in the extraction of typical patterns from a collection of specialised texts, basically noun-noun, adjective-noun or noun-preposition-noun after automatically parsing[12] the text. The parsing process gives greater relevance to those elements which are highlighted by any means (underlining, bold types, etc.).

The second one consists in the application of several filters. Domain relevance is one of them. It is an entropy-based[13] measure which takes into account a candidate's frequency in the specialised domain by comparison with other domains. Domain Consensus (introduced by the authors in Navigli & Velardi, 2002) is also entropy-related and "simulates the consensus that a term must gain in a community before being considered a relevant domain term". Lexical cohesion is another parameter affecting term extraction. The authors follow Park et *al*.'s model (in Sclano and Velardi, 2007: 3) which measures the degree of unithood amongst the constituents of a given pattern. Finally, they employ a set of measures to filter the results with the aim of minimising noise levels (removal of generic modifiers and proper nouns, mispelling detection, etc.).

---

[12] The texts are tagged syntactically after the sentences in it are analysed.
[13] Shannon's Entropy is the key element of Information Theory and represents a way to measure the information in a message. In Statistics, Entropy measures the disorder of a distribution.

Based on all these steps, a word's weight is defined according to Sclano and Velardi "as a linear combination of the three main filters" (Sclano & Velardi, 2007: 3). Let *t* be the CT in question, D_i the domain of interest, *DR* the domain relevance filter, *DC* domain consensus and *LC* lexical cohesion. "The coefficients are user-adjustable, but the default is $\propto = \beta = \gamma = \frac{1}{3}$" (Sclano & Velardi, 2007: 3).

$$w(t, D_i) = \propto \cdot DR + \beta \cdot DC + \gamma \cdot LC^{14}$$

3.2.2.4. Textract (Park et al. 2002)

Park et *al*. (2002) design a term recognition tool, *Textract,* capable of identifying specialised terms which, in their view, convey a major part of the technical knowledge contained in specialised document collections. Moreover, these terms are of great relevance since they can be employed by different applications providing information on syntactic patterns, definitions of concepts or even "relationships that link concepts" (Park et *al.,* 2002: 1).

Term lists can be organised in specialised glossaries, which is the authors' main objective. Glossary formation follows different steps. On the one hand, the identification of CTs (this is the method that will be evaluated herein), on the other hand, the validation of the list of CTs by an expert through its presentation employing a glossary administration system. After that, the validated glossary "is made available, through suitable APIs[15], to the application system" (Park et *al.,* 2002: 1).

Let us then concentrate on *Textract*, the ATR method presented by the authors to single out the most relevant terms in a specialised corpus. *Textract* is part of a set of text

---

[14] For more information on how to calculate the value for each filter see Park et *al*., 2002: 2-3.
[15] Application Programming Interfaces

analysis tools, *TALENT* (*Text Analysis and Language Engineering Technology*), designed by the Information Retrieval and Analysis Group at IBM.

This tool identifies both single and multi-word terms (both noun and verb phrases). The authors apply several filters. To start with, patterns with more than six units are eliminated, proper nouns (person and place names) are removed as well as special tokens such as URLs, words with special characters, etc. Generic premodifiers are also detected, by automatically identifying their level of specificity within a given domain, and purged.

Subsequently, CTs are ranked according to their goodness, that is, their termhood level. Goodness is measured on the basis of a candidate's domain-specificity and the level of cohesion amongst its contituents. The level of specificity of a term (labelled as *confidence* by the authors) is defined as:

$$C(T) = \alpha * TD(T) + \beta * TC(T)^{16}$$

where *TD* stands for term domain-specificity, *TC* for the term's cohesion, and α and β "are constant values which decide the relative contributions of *TD* and *TD* respectively" (Park et *al.,* 2002: 5).

Concerning the evaluation of *Textract,* it is carried out both mechanically and resorting to the help of three judges. Human validation turns out to be more successful as the specialists confirm that 216 (72%) amongst the top 300 candidates extracted are TTs. As for automatic validation, the authors establish the level of overlap between the CTs extracted by their tool and two well-known measures: Church and Hank's (1990) mutual information and Dunning's (1993) log-likelihood. The results of this comparison yield 17.55% overlap for the former as opposed to 55.33% for the latter.

---

[16] For more details on the calculation of *TD* and *TC,* see Part et *al.,* 2002: 5.

3.2.2.5. TermoStat (Droiun, 2003)

As stated in the previous section when describing *Termostat,* Drouin's method can be configured to recognise both SWTs and MWTs. For this section, the parameters were adjusted so that it only extracted noun and adjective phrases, which is the type of MWT pattern this method concentrates on.

The results obtained by the author after evaluating MWT extraction are poorer than those obtained in SWT recognition. While *Termostat* manages to detect 81% SWTs on average, it fails to detect 35% MWTs. As a solution to solve this problem, Drouin points at the possibility of resorting to other types of statistical measures like mutual information (Church & Hanks, 1990), or termhood-weighting factor (Frantzi & Ananiadou's 1997; Nagakawa & Mori's, 2002).

3.2.3. Method implementation

As far as the actual implementation of these five methods is concerned, three of them could be applied in a fully automatic manner. Both Nazar and Cabré's (2012) *Terminus 2.0*, *Termextractor* (Sclano & Velardi, 2007) and *TermoStat* (2003) are freely accessible online. Therefore, the corpus was uploaded and processed automatically producing a list of both single and multiword terms for the implementation of the three methods. Nevertheless, Nazar and Cabré's method required a previous step to the actual processing of the corpus. As shown in the method description section, *Terminus* offers the possibility of training the system so that it can learn what specialised terms are like in every sublanguage. In order to do so, a list of both SWTs and MWTs was uploaded to the server so that *Terminus* implemented the learning algorithm on this data set to improve the term extraction results. This is precisely one of the most outstanding features of this system since the training phase allows it to store a statistical model that

it will apply in the term extraction phase. This information will be saved and made freely available so that any other users willing to process a corpus belonging to the same domain will be able to apply it without any difficulty.

Concerning *Textract* (Sclano & Velardi, 2007) and *C-value* (Frantzi & Ananiadou, 1999), they were implemented using Zhang's (2008) java tool set *Jate. Jate Tools* can be freely downloaded[17] providing the possibility of applying different ATR methods automatically. Zhang's tool set lemmatises and POS-tags corpora using Schmid's *Tree Tagger* (1994). The corpus used for comparison is *BNC*. Zhang resorts to Kilgariff's lemmatised lists, also available online.

However, this tool does not employ any previous filter but rather processes the corpus directly so the output lists resulting after applying *Textract* and *C-value* were filtered employing the function word and base word list 15 of proper names provided with Heatley and Nation's (1996) *Range* software.

3.2.4. Results

The methods evaluated produced different output lists whose size varied depending on the configuration of the different parameters available for each of them. Owing to the need to establish a similar method of comparison, the number of candidates evaluated in all cases was 1400 due to the fact that *Termextractor* by Sclano and Velardi (2007) established a cut-off point producing a maximum amount of 1400 CTs. In spite of the bigger size of the output lists generated by *Terminus* or *C-value,* this was the limit set for the validation of the six methods assessed in this section.

It must be highlighted that the five lists had to be supervised manually once the automatic comparison made with the specialised glossary was finished with the purpose

---

[17] At: http://code.google.com/p/jatetoolkit

of minimising silence throughout this evaluation process. As a matter of fact, those MWTs not present in the glossary were also incorporated to it in the same way as we did with SWTs.
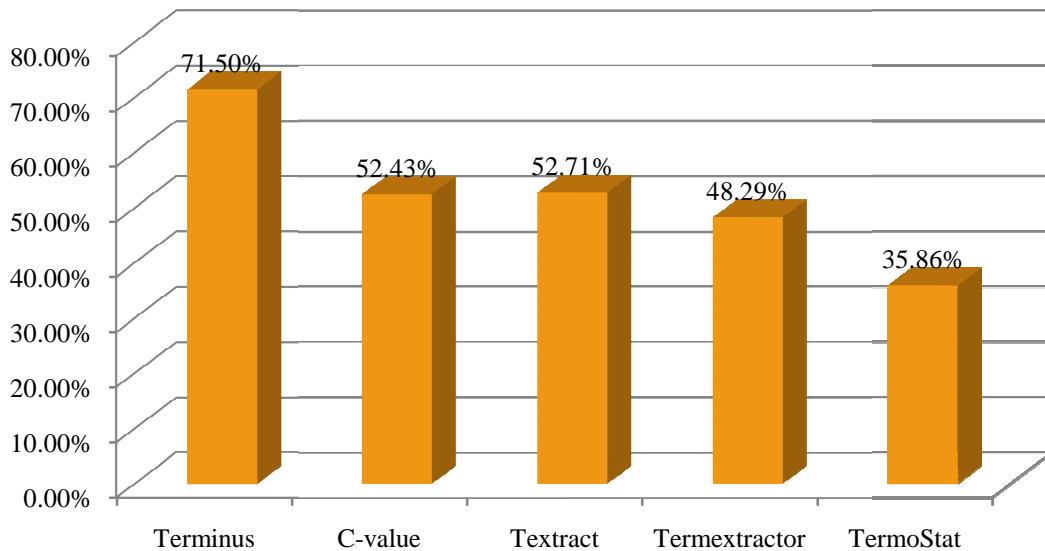
Figure 8. *Average precision of MWT recognition methods*

As far as average precision is concerned for the five methods assessed, figure 8 clearly shows how *Terminus* is the best performing method. It stands 19 points above the second ranking method, *Textract,* which reaches 52.71% precision. C-value is in third position at 52.43% being closely followed by Termextractor, at 4 points below. *Termostat* is the worst performing MWT recognition method which only manages to identify 35.86% MWTs in the corpus. As acknowledged by the author, the efficiency of his method (the most efficient one in SWT recognition), when configured to also detect MWTs, is much lower moving from 85% (its average precision rate in SWT recognition) to 65% (when including both types of lexical units). Nonetheless, for this section, *TermoStat* was configured to only identify MWTs, hence its lower precision.
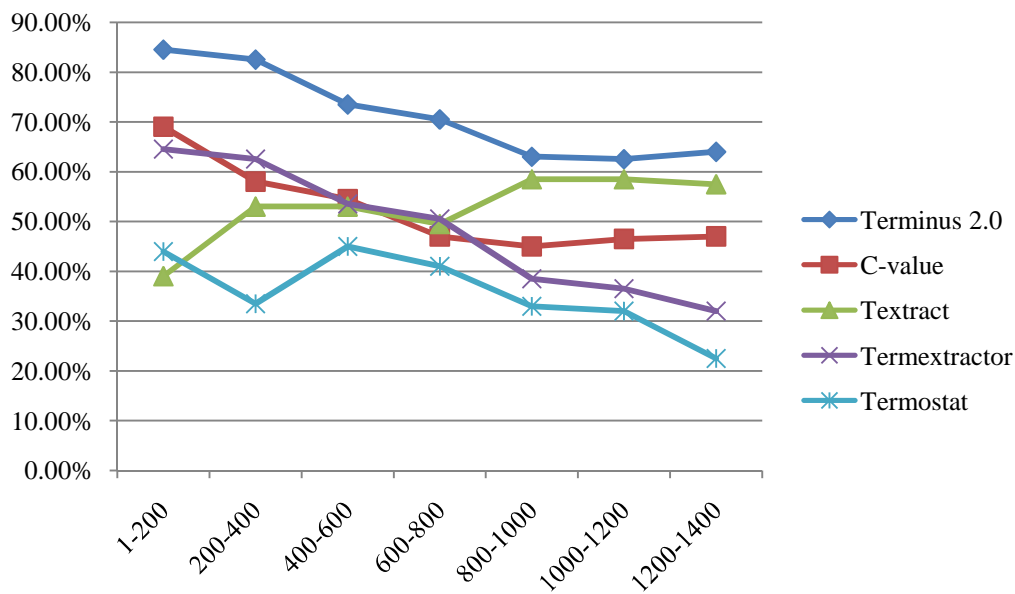
Figure 9. *Cumulative precision of MWT recognition methods for top 1400 CTs*

As relvealed by figure 9, the five methods considered for evaluation behave differently and, although their efficiency is, in general, lower than the one achieved in SWT identification, they do not seem to reduce it as sharply as SWT recognition methods. While the precision levels reached by the latter go down 28 points on average from candidates 1 to 1400 (except for Drouin's method which varies slightly from 84% to 82%, and Chung's whose performance is really poor from the beginning of the CT list), those identifying MWTs do it in a smoother manner.

On average, except for *Terminus* (which goes down 20 points)*,* MWTs recognition methods decrease their efficiency by 24 points within the same range (from candidate 1 to 1400) following a similar trend as shown in the graph above. Conversely, as it happens solely with *RIDF* in SWT extraction (possibly due to the amount of noise detected within the group of 1 to 200 CTs −72%−), *Textract* improves its performance increasing its precision by 18 points although it does not manage to identify more than 57% TTs within the 1400 CT frame studied in this section.

Precisely due to the excellent results obtained on the top 400 CTs (83.5% precision), *Terminus* falls down by 9% from CTs 400 to 600 and continues to descend progressively from that point to the end of the graph still remaining in the first position at the end of it (64% precision).

From CTs 1 to 500, the best ranking methods are *Terminus* and *C-value*, although the latter stands 24 points below the former within this range. *Termextractor* remains in third position from candidate 1 to 700 decreasing its effectiveness from that point to the end of the graph and moving to fourth position from that point on. Finally, *TermoStat* is the worst performing of the five methods evaluated owing to its initial configuration which excludes SWT detection.

Taking into consideration the results obtained in this evaluation process of SWT and MWT recognition methods, it has been proved that the former are more efficient than those which extract either SWTs and MWTs or just MWTs. As a matter of fact, except for *Terminus,* which behaves similarly to *Termostat* within the top 600 candidates in the list, the rest of them are far below SWT extraction methods. Actually, the second best performing MWT recognition method is 14 points below the one in the same ranking position in SWT identification.

Secondly, as far as corpus comparison is concerned, while it yields better results in SWT recognition, it cannot be concluded that it affects MWT recognition positively as three of the five methods assessed above which resort to it rank first, third and fifth respectively. Moreover, it cannot be affirmed either that the greater rate of success of *Terminus* is directly related to the comparison of a general and a specialised corpus but rather to the fact that the system learns about specialised terms when trained by the user

being much more efficient in their identification than others which do not implement any learning algorithm.

Finally, all the MWT recognition methods examined above employ lemmatisation and POS tagging techniques due to the fact that grammatical patterns need to be identified prior to MWT recognition. Therefore, unlike SWT recognition methods where lemmatisation produces better results, it cannot be considered as a relevant factor affecting precision since all the methods studied in this section resort to it.

3.2.5 Processing of *BLaRC*: identification of single and multi-word terms

After having evaluated single and multi-word term recognition methods, *Terminus,* the ATR method designed by Nazar and Cabré (2012), has proved to be the best performing one which manages to identify 71.5% terms in *UKSCC*, achieving 83.5% precision on the top 400 candidates. Therefore, it was employed to analyse *BLaRC*, the 8.85 million-word legal corpus, following a similar procedure to SWT identification in the previous section.

In order to minimise the amount of noise generated by the method, the output list of 5000 CTs was manually supervised to ensure that the automatic validation process had worked properly, two specialised dictionaries (Alcaraz & Hughes, 2000; Saint Dahl, 1999) were employed for such supervision. Those candidates which were confirmed as TTs but did not appear in the glossary and had thus been discarded were added to the gold standard and therefore confirmed as terms. Consequently, the silence generated by the automatic comparison with the glossary was also kept to a minimum.

This manual supervision also led to the elimination of repeated words. *Terminus* lemmatises types not assigning a given weight to each lemma but to its variants. It

includes the different forms of a lemma separately in the output list (indicating the lemma they are associated with) in spite of such forms often belonging to the same morphological category. This might be a problem area for this method which could possibly increase its efficiency if lemmas were considered as single units and their variants were not assigned different weight depending on their forms. An example of this shortcoming is the word *landowner* whose weight in singular is 3576.60925 and 2185.525021 in plural (*landowners*). Hence, the variants of the same lemma were removed from the list[18] to assess precision leading to the elimination of 671 word forms from the original CT list.
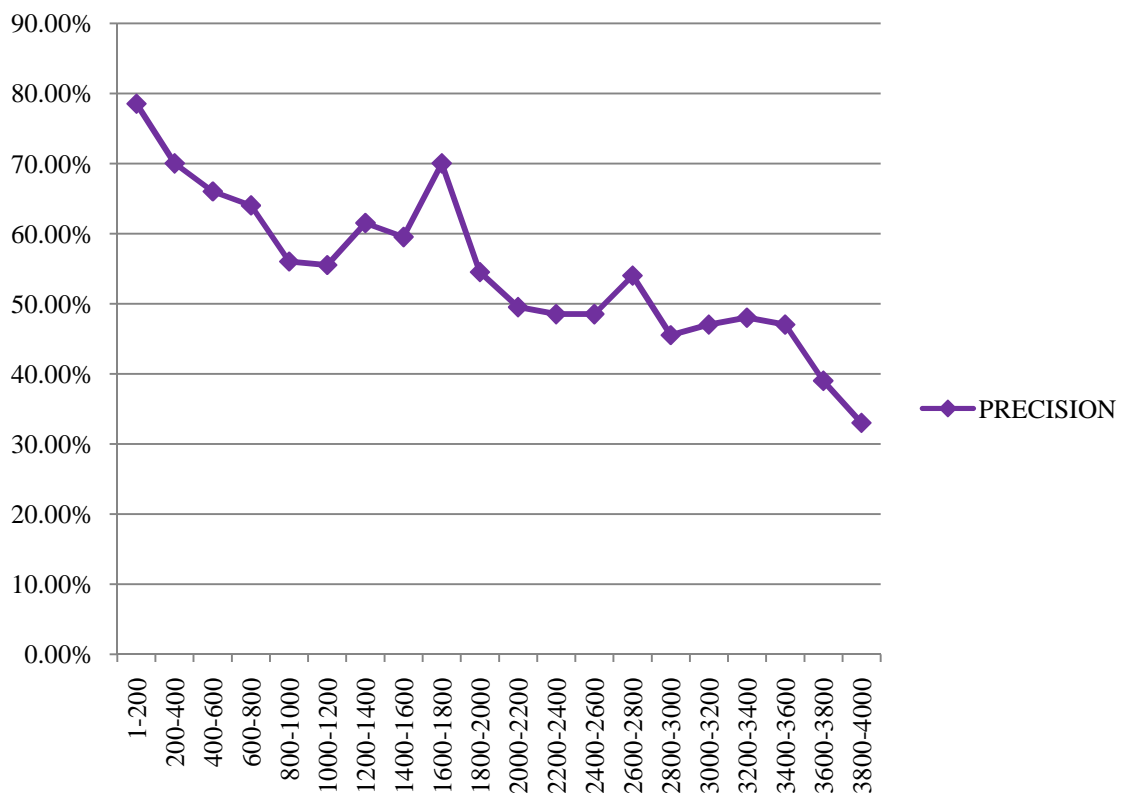
Figure 10. *Cumulative precision obtained by Terminus 2.0 after applying it on*

BLaRC

---

[18] They were removed on condition that they belonged to the same word category like *landowner/ landowners.*

Figure 10 shows the results of the validation process after comparing the whole list of CTs with the gold standard automatically and also supervising it manually. The graph illustrates cumulative precision in groups of 200 candidates from items 1 to 4000 ranked according to the weight assigned to each of them by *Terminus*.

Nazar and Cabré's method does not establish a threshold to discriminate terms from non-terms as clearly as other methods like Drouin's (2003) or Chung's (2003). However, it can be configured so that the number of candidates adjusts to the preferences of the user. In this case, it was configured to produce 5000 terms so the graph above illustrates the evaluation of the first 4000 candidates once the repeated word forms had been eliminated, as already stated.

Terminus remains considerably efficient (especially considering the precision levels achieved by the other MWT recognition methods assessed above) from CTs 1 to 1800, managing to identify 64.5% single and multi-word terms on average within this range and finding its peak at 78.5% for the top 200. Its effectiveness decreases progressively recovering again from CTs 1600 to 1800 at 70% precision. From that point on, it falls sharply to 54.5% and continues to descend smoothly to 48.5% (CTs 2400 to 2600) slightly recovering from candidates 2600 to 2800 (54% precision) and finally falling to 33% by the end of the graph.

Having observed the evolution of this method, it might be interesting to try and establish a cut-off point which would act as a threshold to discriminate terms from non-terms. Judging by the figures, it appears that the method is considerably efficient up to candidates 1600 to 1800 since, after that point, it does not manage to recognise more than 46.77% terms and its precision level decreases rapidly to 33% from candidates 3800 to 4000.

The weight corresponding to CT 1800 is 1030 and could thus be regarded as the threshold value. Applying this threshold, *Terminus* could extract 1153 TTs reaching 65% precision on average.
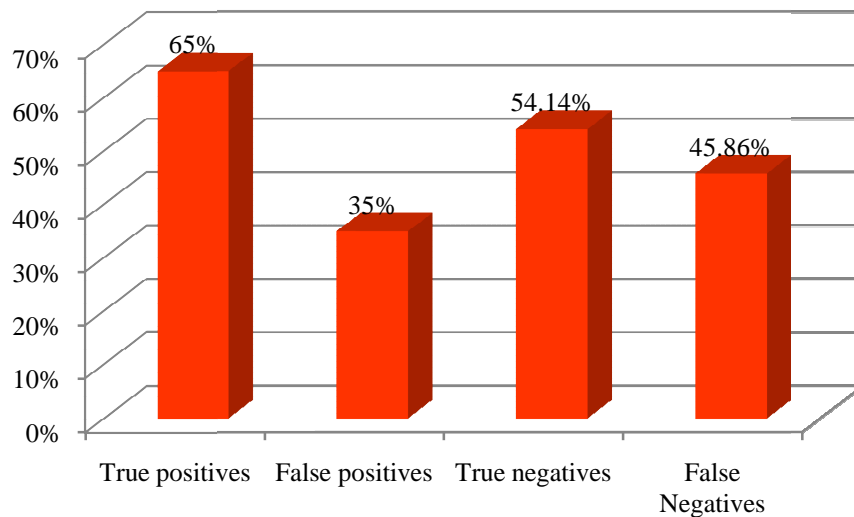
Figure 11. *True and false positives, true and false negatives after applying a + 1030 weight threshold*

As illustrated by figure 11, the percentage of TTs extracted would reach 65%, while noise levels would stand at 35% (percentage of false positives generated by the system). Conversely, establishing a threshold would affect the amount of false negatives, that is, silence, since it would fail to identify 45.86%.

Finally, owing to the fact that it was not possible to have a definite list of MWTs extracted from *BLaRC* to use as reference to calculate recall, partial recall could only be established if we considered the whole list of terms generated by *Terminus* itself, that is, 2312. If the + 1030 weight threshold was applied, *Terminus* would achieve 49.87% recall with respect to the whole list generated by the system without establishing any cut-off point.

To conclude, Table 4 shows the whole list of 2309 terms identified by *Terminus* after its automatic and manual validation without establishing any cut-off point. This list comprises those terms identified solely by *Terminus* as well as the SWTs which coincided with the ones extracted by Drouin's *TermoStat* which were eliminated from the list shown in table 3.

Table 4

List of SWTs and MWTs generated by Terminus 2.0

LEMMA	WEIGHT		
LAWFUL	235189.5609	OBITER	38012.60145
WITNESS	230170.3331	SATISFY	37586.17542
PAYABLE	145321.5141	INTENTION	37040.12247
CAUSATION	135029.1021	PROBABILITY	36475.51462
INJUNCTION	121506.2169	OMISSION	35957.66763
COMPLAINT	112844.924	UNABLE	35472.35907
OBLIGATION	112659.4599	ENACTMENT	35298.79863
INFRINGEMENT	101451.544	RELEVANCE	33939.67726
WORDING	93573.26866	INTERFERENCE	33407.42262
PRESUMPTION	89657.53631	REMISSION	32520.17725
INFERENCE	88221.7819	COMITY	31900.72382
LAWFULNESS	85915.14383	REASONABLE EXCUSE	30386.91454
MISCONDUCT	84649.52406	CONJUNCTION	29289.85025
JUDGMENT	55907.31426	UNDUE	28889.00841
DOCTRINE	54505.99677	PASSAGE	28716.19674
EASEMENT	52735.42997	INADMISSIBLE	28469.58051
INABILITY	51130.07003	ARGUABLE	27817.73845
SUBMIT	47155.40596	ARBITRATOR	27755.13262
IMPUTATION	46254.77462	PRIMA FACIE	27447.29775
CONSEQUENCE	46195.90151	LIABLE	27385.63601
ASCERTAIN	46150.22011	DICTUM	27216.97757
AVERMENT	45711.36982	ENTITY	26378.04993
IMPRISONMENT	43109.06021	EXPIRY	26262.89856
REMIT	42990.04366	ARGUMENT	26106.56628
TENEMENT	41041.95604	PERSECUTION	25897.00009
SPOUSE	40954.20996	CONTENTION	25719.23049
FAILURE	40530.22722	JURISPRUDENCE	25480.94853
ADDUCE	40283.20574	SIGNIFICANCE	25422.11146
ENTITLEMENT	38934.16097	JUSTIFICATION	24554.39017
		DICTA	23849.24221

LEGATEE	23513.67649	STATUTORY DUTY	15197.18437
LOCUS	23196.29014	APPARENT	15063.9798
ADJOURNMENT	22904.33234	COLLUSION	14883.58545
ESSENCE	22826.1898	IMPLICATION	14743.02036
PURSUIT	22138.83194	IMPUGN	14569.62751
PLEADING	21974.17642	ABILITY	14230.74804
CULPABILITY	21286.59623	UNREASONABLE	14217.42858
ERR	21087.46303	RECOURSE	14202.33275
PURSUE	20695.80824	ASSERTION	14118.67235
AMEND	20671.88325	REBUTTAL	14033.86246
ASSUMPTION	20593.27547	CLAIM	13890.29855
COMMITTAL	20369.38632	INAPPROPRIATE	13709.76244
CIRCUMSTANCE	20193.28197	JUSTIFY	13628.56621
CONSIDER	19943.03506	COGNISANCE	13621.25365
PROCEED	19860.84841	PREROGATIVE	13423.13447
INDICTMENT	19471.88392	APPLY	13329.44794
CERTAINTY	18938.07192	ABSCOND	13322.30923
PURSUANCE	18884.94818	INFER	13285.27003
ANNUITY	18663.69979	PREMISE	13094.27684
SEISIN	18367.64791	ADMISSIBLE	13057.17019
ASSESS	18270.01949	CONSTITUTE	13033.94302
REMITTAL	18186.93552	MALICE	13009.98187
ADVERSE	18180.89933	LEGALITY	13002.387
DISHONESTY	18145.69786	INQUEST	12781.68508
DISADVANTAGE	17932.71083	OBSERVATION	12780.31571
ENJOYMENT	17877.16977	EVASION	12778.06256
SEEK	17618.72636	EXPENDITURE	12775.4162
CESSATION	17455.0513	CULPABLE	12699.31371
ARREARS	17124.42591	DECIDE	12543.5875
REDACTION	16957.45434	EXCLUDE	12442.4198
CONCLUDE	16723.35866	ADEQUACY	12432.37187
CONFER	16355.34495	DEROGATE	12347.58389
OVERPAYMENT	16347.31709	TRESPASSER	12325.33619
IRRELEVANT	16319.11289	TRIBUNAL	12241.21948
CUMULATIVE	16253.01775	DURESS	12184.18922
JURISDICTION	16056.04399	IMPARTIALITY	12147.60674
VEXATIOUS	15874.622	CALENDAR	11911.40298
ARBITRATION CLAUSE	15834.45721	CONTEMPLATION	11869.34867
CAVEAT	15775.56498	INHABITANT	11866.19724
REASONABLE TIME	15735.05701	ERRONEOUS	11692.20519
IMPAIRMENT	15578.96533	SUSPICION	11441.87738
PRECLUDE	15571.60762	PAYEE	11437.62708
RESTRICTION	15415.18441	CRIMINAL OFFENCE	11352.19014
QUALIFY	15405.71422	EXPRESS	11309.94326
ENTITLE	15212.21913	MISFEASANCE	11303.62167

93

INDICATION	10987.32092	REJECT	8741.882553
ACCRUE	10986.76424	INHERENT	8592.998279
INCONSISTENT	10631.6676	DETAINEE	8559.713855
IRREDUCIBLE	10626.42803	APPEARANCE	8415.040621
DETERMINE	10548.66808	EX TURPI	8381.394618
VIOLATION	10536.87534	NULLITY	8367.778388
CHAIN	10506.95719	PROVE	8356.465153
UNQUALIFIED	10485.22508	IMMATERIAL	8340.843231
INJUSTICE	10481.48319	CONTEND	8266.946424
EFFECTIVE	10228.27924	CREDIBILITY	8216.235997
CUSTODIAL	10169.95211	COMPLY	8155.083041
PLEAD	10141.45381	REFUSE	8137.667124
ARGUE	10130.73471	NUPTIAL AGREEMENT	8054.335883
ABSOLUTE	10115.75281	INCONSISTENCY	7893.884107
SOLATIUM	10061.7867	PERSONAL DATUM	7880.46792
DISPROPORTIONATE	10006.39646	REASONABLE DOUBT	7741.812898
CONFORMITY	9995.189667	REPUDIATION	7617.533695
ACCRUAL	9983.755174	CONVICTION	7539.657385
REDEMPTION	9974.520703	CAPABLE	7510.947034
MISDIRECTION	9939.227439	TORTFEASOR	7437.347728
JURISDICTION CLAUSE	9903.840005	LAWFULLY	7387.042228
UNLAWFUL DISCRIMINATION	9782.755472	SHAREHOLDER	7314.956715
CONCURRENT	9781.372031	PROPORTION	7288.047433
INTERFERE	9729.847197	DISALLOWANCE	7264.79791
ASCERTAINMENT	9703.129764	FIXTURE	7235.630525
DENIAL	9662.448565	PROROGATION	7192.904206
AMBIGUITY	9630.75077	PEREMPTORY	7179.530454
CAUSATIVE	9629.090509	DUE COURSE	7173.686885
PERSUADE	9584.124739	UNLAWFUL ACT	7153.623464
EXEMPT	9563.857175	CONSIGNEE	7130.001174
RELY	9483.406146	ONUS	7124.171469
NEGLECT	9476.625172	SIGNIFICANT	7121.412092
HEREDITAMENT	9355.16657	MONIES	7054.562892
DISCRIMINATOR	9169.941848	FULFIL	7039.383566
GUILT	9150.95333	INDEMNIFY	7007.328045
IMPOSITION	9136.52734	POSSESSION ORDER	7006.53817
STATUTORY PROCEDURE	9126.142578	TESTAMENTARY	7005.020567
RENUNCIATION	9073.291148	COGNIZANCE	6952.597602
NOTARY	9003.557439	ANTECEDENT	6926.379656
DOUBT	8969.782826	CONFESSION	6921.594246
DISCHARGE	8954.219105	DENY	6917.010697
TESTATOR	8921.177807	EXCLUSIVE JURISDICTION	6840.179787
ATTRIBUTION	8870.669195	GRAVITY	6824.93986
IMPUTE	8751.069492	STATUTORY PROVISION	6815.511133

INTEGRITY	6800.634167	APPRECIATION	5732.7393
IMPOSE	6799.182595	RESIDUE	5709.124845
SEISE	6767.15794	EXCEPTIONAL CIRCUMSTANCE	5707.784906
PROVISO	6764.398186	CONSTRUE	5675.016845
REASONABLENESS	6745.153213	INCAPABLE	5632.523869
MISBEHAVIOUR	6705.429604	EXPERTISE	5616.479935
JEOPARDY	6671.095099	REPAY	5573.475407
DUE DILIGENCE	6666.709751	CONSISTENCY	5570.876582
WITNESS BOX	6652.736039	CONDEMNATION	5566.829417
CONTRAVENTION	6639.148874	ENABLE	5548.785197
ADJACENT	6565.423918	WRONGFUL	5523.979767
CONSTRUCTIVE TRUST	6538.025433	EMOLUMENT	5508.325653
OUTGOINGS	6537.614609	BODILY HARM	5461.357788
BREACH	6528.64733	PERPETRATOR	5428.857445
FORESEEABLE	6514.181557	DISCLOSE	5392.645375
INCOMPATIBILITY	6436.541343	CRIMINAL ACT	5375.458581
NEW EVIDENCE	6406.640402	COMMIT	5359.171362
COHABIT	6354.992955	FORENSIC EVIDENCE	5324.87409
EXPULSION	6326.89382	OCCUPIER	5313.692265
CONVICT	6265.402531	ENFORCE	5312.091084
DISCLAIMER	6251.735216	ADOPTER	5305.986089
ATTRIBUTABLE	6234.338222	COMBINATION	5249.608305
INDETERMINATE	6229.97861	ABANDONMENT	5177.550532
INTERLOCUTORY	6123.467969	INTEND	5169.697694
REASONABLE GROUND	6100.378699	INHERENT JURISDICTION	5130.804542
PAYER	6098.752777	EXCULPATORY	5103.612623
INCRIMINATION	6078.572776	TRIVIAL	5088.788013
EXAMINE	6068.38275	COMPLAINANT	5050.731321
BURGLARY	6057.675494	DEVOLUTION	5048.822946
PROSECUTE	6055.726377	ABOLITION	5043.204573
DEMONSTRATE	6045.049925	VESTED	5023.796883
APPLICABILITY	5989.032996	INCLOSURE	5022.073926
WRONGDOER	5942.830562	UNENFORCEABLE	4994.643358
PRACTICABLE	5919.439325	SEX DISCRIMINATION	4965.99221
RELEVANT EVIDENCE	5908.934355	INCUR	4926.01921
SUPRA	5896.445697	ADJUDICATE	4912.89871
UNPERSUASIVE	5878.284134	RETRIAL	4905.494647
DIVIDEND	5847.912509	ENFORCEABLE	4898.393479
PECUNIARY	5804.362381	UNSATISFACTORY	4883.900369
REASONABLE PERSON	5803.713557	SEQUESTRATION	4883.765391
ACT	5791.315257	ANIMUS	4829.762394
FALLACY	5789.418906	AVER	4807.016928
EJECTION	5776.494457	AUDITOR	4796.617008
CONVEYANCE	5738.503837	APPELLANT	4784.945565

CRIMINALITY	4760.343035	RESCISSION	4274.492294
INCOMPATIBLE	4758.286615	CONSIGNOR	4268.499698
DISHONEST	4757.584424	SERVITUDE	4143.586722
LAWFUL AUTHORITY	4748.379391	CLAIMANT	4124.924169
RESIDUARY LEGATEE	4738.883393	ASCERTAINABLE	4101.491829
PARTICULAR CASE	4734.67674	LEGITIMACY	4090.253426
AGGRAVATING	4708.588468	REASONABLE SUSPICION	4083.137416
INJURE	4684.123577	DEDUCT	4075.079908
LANDOWNER	4681.508841	SUPERSESSION	4057.844379
FLAW	4678.179783	QUASH	4055.161634
CUSTODIAL SENTENCE	4642.349308	WILFUL	4042.25625
INTERVENE	4642.10532	CONTROLEE	4041.147365
DISALLOW	4642.105299	REASONABLE PROSPECT	4039.842942
UNFIT	4632.607571	DISCRIMINATORY	4033.382957
EXPEDITION	4616.0838	UPHOLD	4000.709291
LACHES	4613.010194	DONEE	4000.233184
WHIPLASH	4579.019128	EXCESSIVE	3981.296297
IRREGULARITY	4557.604384	ESTOPPEL	3965.990976
PROBABLE	4556.762501	MARGIN	3947.666083
DEPORTEE	4554.946698	FIDUCIARY DUTY	3942.523707
DRAFTSMAN	4553.844469	AXIOMATIC	3938.804545
REVOKE	4548.042276	EXTANT	3922.997461
ALTERATION	4546.818012	INFLICT	3916.133556
DESIRABILITY	4542.748228	TORTIOUS	3902.316702
PROPRIETY	4537.678429	ONEROUS	3893.84995
DPA	4536.788745	OBJECTION	3877.385776
SUSPEND	4521.228481	STATUTORY TEST	3874.710516
REDEVELOPMENT	4521.113007	FULFILMENT	3857.070997
BENEFICIAL	4507.332956	JUSTICE	3825.411071
REASONABLE ADJUSTMENT	4489.413752	LAWFUL SPORT	3821.528006
DEAL	4483.446913	PROPORTIONATE	3817.738208
DEMISE	4478.122402	DEFAMATORY	3797.678666
DISREGARD	4471.496382	BUGGERY	3793.422716
DESCENT	4404.219868	ACCEDE	3781.159131
EARNING	4386.904585	HOMICIDE	3760.173898
ACCUSATION	4383.088966	INFRINGE	3740.973007
STRICT LIABILITY	4373.982624	INCIDENT	3731.231617
BAD FAITH	4360.439974	ACCORDINGLY	3728.696117
REJECTION	4346.309094	BILATERAL	3728.674999
DEPRIVE	4345.584234	INDEBTEDNESS	3719.880451
FRESH EVIDENCE	4339.097313	WITHDRAW	3714.523108
SUBLET	4324.746243	THREATS	3714.364426
HIRER	4308.479185	LEGITIMATE AIM	3710.526871
INTERVENTION	4300.83591	TENEMENT	3706.940935

FAVOURABLE	3692.178095	ADJOURN	3231.355338
MITIGATE	3686.679733	CONFINEMENT	3227.448451
PCT	3668.634526	SUCCESSOR	3205.167762
REBUT	3659.890053	HARMFUL	3198.212328
APPEAL	3657.253122	PERMISSIBLE	3197.89955
EVICT	3656.584758	APPARATUS	3190.297371
INTIMIDATION	3650.438445	CONTROVERSY	3189.602762
RESPONSIBILITY	3631.386502	BREACH OF CLAUSE	3187.081327
PUTATIVE	3612.508234	ACCREDITATION	3183.369187
DONOR	3606.679468	ACCEPTABLE	3172.22741
CONTRACT OF EMPLOYMENT	3592.032606	COMMON GROUND	3168.48596
ARGUABLE CASE	3585.730212	PREVAIL	3162.797677
UNMARRIED	3572.202025	MESNE	3153.355992
JURISDICTIONAL	3569.467934	LIMITATION PERIOD	3153.315967
INDORSEMENT	3561.433735	LESSER	3148.592214
BEQUEST	3557.199383	CONSENSUS	3143.891661
CRIMINAL LIABILITY	3546.017563	INDEMNITY CLAUSE	3143.516067
PARTICIPANT	3544.524714	DISTRIBUTOR	3136.64045
BURDEN OF PROOF	3515.317722	LEGACY	3127.269122
CONFIRM	3507.659506	ABETTOR	3126.67162
BAILEE	3503.484313	UNJUSTIFIED	3120.853053
STATUTORY ASSUMPTION	3501.448013	STATUTORY DEMAND	3117.360659
IMPRACTICABLE	3496.643878	JUDICIAL DECISION	3115.358659
ENDORSEMENT	3464.858482	ARGUABLY	3115.137014
TERMINATE	3461.950814	SUFFICIENT EVIDENCE	3100.33273
EX TURPI CAUSA	3417.552236	CRIMINAL PROPERTY	3096.443363
INADMISSIBILITY	3388.374723	INCRIMINATE	3084.194654
MANUSCRIPT	3383.476117	ALLEGE	3071.577629
IMPARTIAL	3361.86066	FACTUAL BASIS	3062.791739
UNLAWFULNESS	3360.080359	MALADMINISTRATION	3060.80002
HABITATION	3357.929424	TRUST	3060.728391
AUTHOR	3357.871959	HOUSEHOLD	3058.120234
PURPORTED	3353.520004	IMMIGRANT	3051.220783
ADVERSE EFFECT	3347.931148	REASONABLE CAUSE	3043.942285
ORDER FOR POSSESSION	3343.400952	UNDESIRABLE	3038.152776
INVOKE	3325.428236	MALICIOUS	3032.32696
REIMBURSEMENT	3293.08982	COERCION	3031.832103
DEFICIENCY	3290.965676	THEREBY	3025.11155
ROBBERY	3266.771413	SPECIAL ADVOCATE	3013.889537
MATRILINEAL	3255.229176	UNARGUABLE	3011.030429
DISSENT	3243.871894	CONCEAL	3010.044078
IMPARTIAL TRIBUNAL	3243.442881	COGENT EVIDENCE	3004.644815
SUBSTITUTION CLAUSE	3233.149795	INFLICTION	2999.737878
		CULPABLE HOMICIDE	2997.820694
		DEFAULT JUDGMENT	2986.108533

PRECONDITION	2985.659764	OBSTRUCTION	2665.757115
SEEKER	2975.480502	COMPARABLE	2650.563667
PARTICULAR CIRCUMSTANCE	2973.037788	HYBRID AUTHORITY	2648.811439
CORRECTNESS	2968.139334	GENUINENESS	2641.670705
INDECENCY	2964.664878	CRIMINAL TRIAL	2636.758439
DEROGANT	2963.110624	INNOCENCE	2616.976971
MAIN JUDGMENT	2951.205661	SPECIAL CIRCUMSTANCE	2616.911408
RELEVANT TIME	2949.680843	ORAL SUBMISSION	2615.995158
CRIMINAL	2922.646613	INTESTACY	2609.191019
FALSEHOOD	2900.427637	LAWFUL DETENTION	2597.729202
STATUTORY CONSTRUCTION	2886.153855	WAIVE	2587.660666
NULLIFICATION	2864.586688	RATEABLE	2578.196095
ABIDE	2857.1408	THEREOF	2570.904393
LEGAL TEST	2854.916164	INSUFFICIENCY	2565.903578
FORFEIT	2840.703923	INADEQUATE	2552.521741
SUFFER	2838.582457	EXTRADITE	2545.721676
EFFLUXION	2829.853593	DETERMINATE SENTENCE	2538.726071
PROFESSIONAL PRIVILEGE	2818.931194	JUDGMENT DEBT	2534.323392
EFFECTIVENESS	2814.295579	MATRIMONIAL HOME	2528.885457
STATUTORY POWER	2811.546076	LAY	2526.497603
IRRECONCILABLE	2805.703956	SUBORDINATE	2525.347885
EVIDENCE	2805.657392	RELEVANT PROVISION	2521.201598
EXTRINSIC EVIDENCE	2801.895582	JUSTIFIABLE	2519.872098
FULL AGREEMENT	2779.796892	INSUFFICIENT EVIDENCE	2513.938995
EXAMINATION	2777.396057	MISDIRECT	2511.799349
MITIGATING	2769.630402	COURT	2503.815875
FRESH CLAIM	2767.602955	NEGOTIATION	2498.537332
LIQUID	2756.67838	LEGITIMATE EXPECTATION	2493.97021
INNUENDO	2755.90498	CHARTERER	2464.043328
CONCUR	2749.069678	CAUSAL CONNECTION	2458.413253
WEDLOCK	2744.714571	PATENTEE	2457.724211
GRIEVOUS	2739.376204	EXPRESSLY	2453.04087
SEGREGATION	2736.183384	CREDIBLE	2450.432804
DETERIORATION	2711.756813	FRAUDULENT EVASION	2449.00851
INCONVENIENCE	2709.708177	EXIGENCY	2443.165186
CONVENIENT	2695.58331	FLAGRANT	2438.835717
ERROR OF LAW	2694.821007	COGENCY	2419.591879
POSSESSION PROCEEDING	2692.895014	CERTIFY	2407.110358
SATISFACTORY	2691.809724	APPREHENSION	2404.198114
ENEMY	2688.877847	REVEAL	2401.390935
VAT REGISTRATION	2685.564202	MISCHIEF	2382.773192
DISTURB	2677.314003	NATURAL JUSTICE	2376.415029
INTEGER	2666.5228	HEARING	2375.8187

OBJECTOR	2373.907461	TESTATRIX	2159.99544
HEADNOTE	2370.851503	PUBLIC DOMAIN	2155.583406
UNLAWFULLY	2370.618382	QBD	2153.44404
SUPERSEDE	2366.542387	DELIBERATE	2152.485512
DENOTE	2361.248192	UNAMENDED	2150.795979
RELEVANT PROPERTY	2358.38777	CONTRAVENE	2147.31918
REVISE	2356.835607	DIRECT EVIDENCE	2140.421791
JUDICIAL DISCRETION	2353.891627	INCOMPETENT	2136.844953
FORESEE	2350.303505	PRIVILEGED	2134.715746
VICARIOUSLY	2349.036288	PROBATIVE	2132.922605
KEEPER	2343.337029	EMPLOYMENT	2121.523287
LEGAL CERTAINTY	2337.269289	CONNEXION	2119.663913
STATUTORY LANGUAGE	2332.567561	RACIAL DISCRIMINATION	2104.817745
EFFICACY	2321.006615	COLLECTIVE ENFRANCHISEMENT	2097.888241
JOINT ENTERPRISE	2312.377204	LEGAL PROCEEDING	2095.985744
INSOLVENT	2298.706389	BALANCE OF PROBABILITY	2094.702104
CRIMINAL STANDARD	2294.91227	CONSIGNMENT	2089.87153
ATTRIBUTE	2286.792326	ASSIGN	2088.241276
INSERTION	2283.716155	REPRESENT	2081.452806
INDETERMINATE SENTENCE	2281.797099	MAINTAIN	2074.063911
STATUTORY DEFINITION	2275.69576	LEGAL PRIVILEGE	2071.340753
EXPERT EVIDENCE	2271.795159	NUGATORY	2070.07451
MISAPPROPRIATION	2264.963288	NEXUS	2068.397616
PROHIBIT	2261.908121	AFFIDAVIT EVIDENCE	2062.337171
MECHANISM	2254.102208	QUANTIFICATION	2058.25439
STATUTORY PURPOSE	2249.815343	PRESCRIPTIVE	2056.18427
TENOR	2246.839571	VINDICATION	2054.745328
EXCLUSIVE COGNISANCE	2242.10522	GRAVAMEN	2050.223626
REMEDIAL	2241.05402	IMPRACTICAL	2048.203266
MENTAL DISORDER	2238.997563	INVESTIGATE	2048.113485
ENACT	2235.568817	ULTERIOR	2046.380157
BEREAVEMENT	2221.62283	FETTER	2042.211399
ADVERSE POSSESSION	2218.649035	EXCEPTIONALITY	2041.146417
ASSERT	2212.223032	JUDICIAL AUTHORITY	2039.708592
UNFETTERED	2206.925889	FAVOURABLE TREATMENT	2025.993611
RELEVANT CLAUSE	2206.310326	CHATTEL	2021.891216
CONCURRENCE	2205.657873	SPECIAL ADJUDICATOR	2019.806079
ACQUIESCENCE	2204.621646	DECISIVE	2017.638106
ANCILLARY RIGHT	2189.010465	UNDUE INFLUENCE	2010.754616
INTEREST OF JUSTICE	2180.781184	ENFRANCHISE	2006.539589
MTIC FRAUD	2180.341186	TESTAMENTARY CAPACITY	2002.701207
CONSUMPTION	2178.221984	RESCIND	2000.866049
EXPRESS EASEMENT	2168.286208		

FREEHOLD	1993.198092	PREJUDICIAL	1822.054353
DOMINANT	1985.512232	PROVOCATION	1808.359175
PROHIBIT	1982.081981	COMPLAIN	1806.701091
DISCRIMINATE	1964.959534	REBUTTABLE	1805.670535
PURVIEW	1964.889639	OMIT	1786.959926
EXECUTE	1957.371287	STATUTORY TIME	1776.643558
RELEVANT AUTHORITY	1953.618876	OPPRESSIVE	1773.68223
STATUTORY CODE	1952.076041	INFIRMITY	1769.37818
PRESCRIBE	1949.368237	EVIDENT	1761.682144
EXCUSABLE	1948.282515	REDEEM	1761.530771
CONSPIRATOR	1942.367885	EXTEMPORE JUDGMENT	1750.759899
UNAMBIGUOUS	1939.93049	CONSCRIPT	1745.471297
CRIMINAL ACT	1939.851808	CONSANGUINITY	1741.008615
MATERIAL TIME	1934.857892	JUDICIAL	1738.93781
KINSHIP	1934.723925	EXERCISE	1738.07524
PERSUASIVE	1930.647062	BAILOR	1737.380755
INVESTIGATOR	1929.407647	SOLEMN	1731.908101
LAWFUL CUSTODY	1926.638077	CONSULT	1731.515783
DISMISSAL PROCEDURE	1925.137601	COMMITMENT	1730.814692
RATEABLE VALUE	1922.921898	MATERIALITY	1725.227042
MOVE	1917.304133	IDENTIFIABLE	1724.266461
REMITTANCE	1914.136692	MENTAL IMPAIRMENT	1723.675504
WITNESS EVIDENCE	1913.2247	FAIR HEARING	1722.369804
CONTRACTOR	1912.147316	INPUT TAX	1717.027877
NOMINAL	1910.92987	ENVISAGE	1716.094219
REALISABLE PROPERTY	1908.009118	UNDULY	1710.920057
LIEN	1906.99208	REGULATOR	1709.731162
COMPULSORY ACQUISITION	1903.423636	UNQUALIFIED RIGHT	1708.953108
EFFECTIVE DATE	1894.171941	STATUTORY OFFENCE	1707.664536
IRREPARABLE	1893.728357	INSURANCE CONTRACT	1707.423518
RESIGNATION	1891.284176	FORESEEABILITY	1700.577169
PROCEDURE	1889.699582	CONVENTION	1700.509621
PERPETUITY	1881.877071	LAWFUL RESIDENCE	1697.260063
INSTRUCT	1880.086862	AMENABLE	1695.268872
SFO	1874.587084	DEFEND	1689.396862
SERIOUS BREACH	1871.14976	CRIMINAL CHARGE	1689.31132
LAW OFFENCE	1857.909251	TERM OF TRUST	1687.505604
REPATRIATION	1854.528749	SUBORDINATION	1677.875788
IRREVOCABLE	1845.339106	UNPERSUADE	1675.361632
PROXY	1843.366679	CRIMINAL JURISDICTION	1672.635798
INCIDENTAL	1834.315062	VOIDABLE	1670.386597
PROSPECTIVE	1831.497642	RETAIN	1666.778091
INTRUSION	1824.394149	PROMISSORY	1659.595744
		PROMISSORY	1658.885049

ESTOPPEL		RELIABLE WITNESS	1565.227095
DETAIN	1656.877677	CONFESS	1564.446433
HAZARDOUS	1655.647647	CONFINE	1564.136528
RSC	1651.498616	JUDICIAL PROCEDURE	1562.610794
REVERSION	1650.242235	DISQUALIFY	1560.822575
UNOBJECTIONABLE	1647.824391	OATH	1557.015065
FIXATION	1647.10831	PRETENCE	1556.216944
POTENTIAL	1645.575633	TRANSLATION	1554.013783
PATERNITY	1642.614295	DIVERGENCE	1552.672852
DAMAGE	1638.909017	ANTERIOR	1546.675761
QUANTUM JUDGMENT	1638.863286	COMPETENT JURISDICTION	1543.051247
RETROSPECTIVITY	1638.443209	PRECEDENCE	1538.344868
VOLITION	1637.993182	UNADMINISTERED ESTATE	1532.658452
DEVOID	1637.977123	ENCROACHMENT	1531.628595
IMPRACTICALITY	1632.165163	LAWFUL JUDGMENT	1530.764552
PROMULGATE	1626.859562	LAWFUL ORDER	1529.862702
CONCEALMENT	1626.756904	TENANT COVENANT	1529.279882
ADJUDICATOR	1621.969539	CUSTODIAL PERIOD	1528.40122
CRIMINAL PROSECUTION	1621.880819	HITHERTO	1528.32749
NEPOTISM	1619.772465	RENDER	1527.862291
ARTICLE	1613.383337	SUBROGATION	1520.349129
PENDENCY	1612.56415	STATUTORY WORDING	1518.408564
LEGAL ASSISTANCE	1606.245281	COERCIVE	1509.807795
RATIONAL	1606.208288	LAWFUL ARREST	1507.890589
COMPETENCY	1604.936225	TRUSTEE IN BANKRUPTCY	1507.774642
LEGAL REPRESENTATIVE	1601.928177	ENDORSE	1502.02449
DEBENTURE	1596.35286	ADDITIONAL EVIDENCE	1500.68729
BREACH OF TRUST	1595.754743	LAWFUL EXCUSE	1496.603109
PERSUASION	1592.911442	OBJECT	1495.449277
ADVERSE INFERENCE	1587.393969	MISAPPLICATION	1490.58424
WRONGFUL ACT	1587.008356	RECKLESS	1489.947073
PRIVATE LAW	1583.849495	PRECAUTION	1489.456905
DIVERSION	1583.252672	LIQUIDITY	1486.219249
REVERSAL	1581.301422	HEREINAFTER	1486.107407
EXTRADITION OFFENCE	1579.449588	LAWFUL MEANS	1485.853794
DEPRECIATION	1577.702082	IMPEDIMENT	1484.768983
DILATORINESS	1577.218419	CONSCIOUSNESS	1483.993629
INALIENABLE	1575.777737	POSSESSION CLAIM	1480.37997
INEFFECTIVE	1575.416934	LEGALLY	1474.721134
HEAR	1573.472005	CONCLUSIVE	1473.515476
SERIOUS INJURY	1570.90851	UNSUITABILITY	1472.805046
FINAL JUDGMENT	1568.104282	WRONGDOING	1469.833302
LEGAL DUTY	1567.643149	PERSONA	1467.988374
EXEQUATUR	1566.155806		

ORAL ARGUMENT	1464.451405	DISCRETIONARY TRUST	1384.922852
ERASURE	1463.784129	PRIVITY	1384.202531
ENTRANT	1458.430922	UNCONSCIONABLE	1382.560898
EXCLUSION CLAUSE	1454.747234	ENFORCEABILITY	1382.27836
AVAIL	1452.104162	MERE FACT	1381.484123
MUTUAL TRUST	1451.044789	RETAINER	1380.902613
SUMMARY DISMISSAL	1451.032599	CRIMINAL ACTIVITY	1380.383241
CLEAR EVIDENCE	1447.654641	JUDICIAL CAPACITY	1379.318916
UNTRUE	1446.187572	INHERIT	1376.69829
GRANTOR	1444.418999	SENIORITY	1374.879543
CUSTODIAL TERM	1441.957642	OVERRIDE	1371.83882
OCCUPANCY	1440.338558	FREEHOLDER	1371.38971
INFLUENCE	1438.67151	RELEVANT CONSIDERATION	1370.402822
RECTIFY	1437.448541	OCCUPANT	1367.179407
AFRESH	1437.44468	INJUNCTIVE	1364.497279
OPEN COURT	1431.766297	ARTIFICIAL	1362.687971
OSTENSIBLE AUTHORITY	1426.546474	LITIGATE	1360.475005
DERIVATIVE CLAIM	1426.452767	PROSPECTUS	1359.342938
UNLICENSED	1425.536191	ANTECEDENT OFFENCE	1357.60894
ARBITRARINESS	1423.511768	EFFECTIVE REMEDY	1357.250033
OBSERVANCE	1422.937673	LIVE EVIDENCE	1356.999396
AGGRIEVE	1422.905337	JUDICIAL BODY	1353.500025
MODIFIED PROCEDURE	1421.35423	PRIMACY	1342.745758
ASYLUM SEEKER	1418.560194	WHEREABOUTS	1341.801622
CERTIFIED	1417.546914	BENEFIT	1340.018954
ENTRUST	1416.946134	FOREGOING	1336.679591
EXTRADITEE	1415.637364	VALID	1334.584915
PREPONDERANCE	1415.530273	UNAPPEALABLE	1330.036853
REPUDIATORY	1414.4476	GOOD EVIDENCE	1329.538045
ACTIONABLE	1411.840024	INADVERTENCE	1323.394697
SUE	1409.594091	MISNOMER	1321.720367
POSSESS	1405.561765	INACCURATE	1320.031908
LEGAL RIGHT	1404.468003	COLLUSIVE	1315.083518
RELEVANT OFFENCE	1403.197321	HERITABLE PROPERTY	1312.663751
DEBT RELIEF	1402.229605	INCLUSIVE	1312.417214
TANGIBLE	1397.863182	PUNITIVE	1311.848291
MISCARRIAGE OF JUSTICE	1394.513463	LEGAL EFFECT	1306.338683
ARBITER	1393.156522	IMMEDIATE RISK	1304.832272
IMPERMISSIBLE	1389.168053	CONTUMACY	1304.124953
CONTINUANCE	1387.87898	EXPIRATION	1302.482161
UNDERTAKER	1387.049559	COMMON LAND	1300.992465
INTERMEDIARY REPRESENTATION	1385.25082	JOINT TENANCY	1299.910002
FACILITATE	1385.005641	TURNOVER	1297.890072

ACQUIT	1297.431165	JOB TITLE	1237.505428
WITHHOLD	1296.917828	LEGISLATE	1236.034226
DETERRENCE	1296.123525	REASSESSMENT	1234.135552
DEPORT	1291.444577	SEVERAL LIABILITY	1232.529559
ENQUIRE	1289.927883	REMUNERATE	1232.421213
RECOMMENDATION	1287.540949	COMPARATOR	1229.605062
JUDGE	1287.467382	DETERMINATIVE	1229.445831
TRAP	1286.620665	FALSE IMPRISONMENT	1229.051181
HOTCHPOT	1286.554461	ADOPTIVE PARENT	1228.016851
CHAIR	1286.504275	EGRESS	1225.351581
CITE	1285.860881	SERIOUS OFFENCE	1224.379435
CRIMINAL RESPONSIBILITY	1284.045497	SUMMARY CONVICTION	1223.068572
GROUNDS OF APPEAL	1283.976503	CONDITIONAL DISCHARGE	1222.80716
INHIBIT	1283.062319	LEGAL PRINCIPLE	1220.550893
AFFINITY	1282.861442	DEVIATION	1216.988271
ABROGATION	1281.097266	EXERCISABLE	1216.652262
LEGAL TITLE	1278.312949	EYE WITNESS	1212.900117
EXPEDIENT	1275.329054	TRIABLE	1206.019253
CARELESSNESS	1274.067433	RESPONDENT	1205.26457
ACCELERATION	1273.752544	CIVIL RIGHT	1204.37631
INTRUDER	1272.648844	STATUTORY FUNCTION	1203.708191
TAINTED	1271.9991	JUDICIAL NOTICE	1202.066163
CREDIBLE WITNESS	1270.820693	DEFAULTER	1201.53002
AIRSPACE	1270.57717	EXEMPLARY	1200.780419
OBVIATE	1267.49327	IMMIGRATION	1200.470121
REFER	1267.464964	BONA FIDE	1198.727705
COMMUNICATE	1266.226439	SUBSTANTIVE JUDGMENT	1197.548202
DEEM	1265.01171	PROMULGATION	1191.97785
VITIATE	1264.950915	COMPEL	1188.305923
REGISTER	1264.394123	SUBSTANTIVE HEARING	1186.615745
STANDARD PROCEDURE	1263.594754	JUDGMENT DEBTOR	1185.779794
LACUNA	1262.707744	JUDICIAL PROCESS	1185.571907
EDUCATION AUTHORITY	1257.089599	DERIVATIVE ACTION	1185.454235
REMAND	1256.760514	PROBATIVE VALUE	1185.389346
PRECAUTIONARY	1255.20431	IMPAIR	1184.536265
REFUTE	1254.729695	CHECK	1183.846452
IRREBUTTABLE PRESUMPTION	1253.838394	INTERIM STATUTE	1182.667145
INTERPRETER	1251.76882	STATUTORY APPEAL	1181.707199
STIPULATION	1249.860417	DISTINCTIVE	1179.216245
TITLE DEED	1249.848177	REBUTTABLE PRESUMPTION	1177.977508
JOINT ADOPTION	1244.122717	FINDING OF FACT	1175.600982
HONEST	1242.74355		
THIEF	1238.366059		

DRAFT	1174.96136	CONSCIENTIOUS	1124.582771
DETRACT	1174.683304	PERMANENT RESIDENCE	1123.032687
JUDICIAL TRIBUNAL	1173.022806	CONTEMPLATE	1123.016173
DESTRUCTION	1171.450622	PRIVY	1121.37142
RESTRICTIVE COVENANT	1170.387592	MATERIAL FACT	1120.664138
RESIDUARY ESTATE	1169.886408	SECRECY	1118.751112
TRUSTEE	1169.344827	ADDUCE EVIDENCE	1118.416589
STATUTORY REMEDY	1168.887052	TRY	1115.759343
REVALUATION	1168.410417	LEGAL POSITION	1115.758872
MENTAL CONDITION	1167.442508	EQUITABLE CHARGE	1114.19283
INGRESS	1166.31041	PROTECTED	1113.430951
COVENANT	1164.002659	EIR	1113.420986
PREPARATORY	1162.898255	ORAL AGREEMENT	1112.815709
CULPA	1162.478406	ONLY WITNESS	1111.688229
ORDINARILY	1160.49831	UNDIVIDED	1110.983975
SOLVENCY	1160.275924	WLR	1108.432331
DECLARATOR	1158.29253	CONVENTION	1108.250053
FAIR PROCEDURE	1158.07445	VAT LIABILITY	1106.820324
INEQUITABLE	1156.192464	TRACE	1106.684261
SENTENCE OF IMPRISONMENT	1155.539915	INAPPLICABLE	1105.629758
TAX LIABILITY	1154.929114	MESNE PROFIT	1104.34322
OPEN JUDGMENT	1152.369329	UNANIMOUS	1104.325195
RENTAL	1150.220839	JOINT TORTFEASOR	1104.257942
ASSIGNEE	1149.875261	INQUISITION	1100.283115
EXTENSION	1149.576831	STATUTORY PROHIBITION	1099.298768
RECONSIDER	1149.359033	ADHERE	1098.985099
SUBJECTED	1148.796889	INADMISSIBLE EVIDENCE	1096.660356
PEREMPTORY CHALLENGE	1147.013449	COGNATE	1096.079538
ANTICIPATE	1144.126793	FEE AGREEMENT	1095.446909
CUSTODIAL BEHAVIOUR	1144.083377	INDEPENDENT TRIBUNAL	1094.605949
INJUNCTIVE RELIEF	1143.479866	PEDIGREE	1094.394679
RECONCILE	1141.47807	OBEDIENCE	1093.312969
PRESERVE	1141.230992	CRIMINAL ENTERPRISE	1092.750882
UNEXCEPTIONABLE	1136.900606	SURVIVORSHIP	1092.346549
FACTUAL MATRIX	1133.080443	NON-DEROGATION	1090.812067
FRAUDULENT TAX	1133.051462	GOVERN	1089.586956
CRIMINAL DAMAGE	1132.913133	CRIMINAL CASE	1089.471548
POTENTIAL LIABILITY	1132.827141	ADMINISTRATIVE PROCEDURE	1089.105323
CONDUCIVE	1132.276481	OPPONENT	1089.023177
ABATEMENT	1130.408669	MARITAL STATUS	1088.029982
ANONYMOUS WITNESS	1126.927092	CANCEL	1087.39708
REASONABLE BELIEF	1126.882413	SEXUAL IDENTITY	1087.291012
INACCURACY	1125.71761		

UNFOUNDED	1085.672218	HOSTILITY	1033.138268
RELEVANT INFORMATION	1084.54612	MANIFEST	1032.953077
FACTUAL CIRCUMSTANCE	1082.922487	ENLARGEMENT	1032.916316
PROFFER	1081.054134	LEG	1032.463831
CONTRADICTION	1079.524101	LEAKAGE	1032.067523
OBJECTIONABLE	1078.256561	AFFORD	1030.521709
LEGAL OBLIGATION	1077.720995	COHABITEE	1028.885561
HIGH AUTHORITY	1075.893617	BENEFICIAL OWNER	1028.098921
APPROPRIATE SENTENCE	1074.148381	APPROPRIATE REMEDY	1024.801411
STATUTORY OBLIGATION	1073.986987	LIFE IMPRISONMENT	1024.281712
PRIOR ART	1072.704241	EVALUATE	1023.263037
ACTION ESTOPPEL	1070.656229	APPURTENANCE	1022.5516
INSTIGATION	1069.64401	SCRIPT	1022.251902
LEGAL AUTHORITY	1069.495629	IMMOVEABLE PROPERTY	1021.406967
RELEVANT PART	1068.39969	CRIMINAL BEHAVIOUR	1019.873398
LORDS	1066.242312	UNWILLING	1018.706189
EXCULPATE	1065.585098	IRREBUTTABLE	1017.175679
DECLARATION OF TRUST	1061.849419	WITNESS INTIMIDATION	1014.241491
NOVATION	1059.781243	ALLOWABLE	1013.53553
ARBITRATION	1059.31085	APPOINTEE	1010.488756
BENEFICIAL INTEREST	1058.644326	MORTGAGEE	1008.445678
RELEVANT MATERIAL	1057.969505	JUDICIAL INTERPRETATION	1008.093951
FEASIBLE	1056.91037	ALIENATION	1002.819226
CONFRONTATION	1056.475087	INIMICAL	1002.388199
TRUST FUND	1053.589501	EFFICIENT	1001.864729
OBITER DICTA	1052.88065	BEHAVE	1000.898414
TRUST PROPERTY	1052.682901	ADMISSIBLE EVIDENCE	998.646684
ARSON	1050.115818	ASSIGNATION	998.102487
IRREDEEMABLE	1047.517683	SEX OFFENDER	997.907661
FACE VALUE	1047.197687	UNOCCUPIED	993.974864
EWCA	1045.55451	ESTIMATION	990.473352
MATERIAL ERROR	1045.407057	GIVE NOTICE	989.846483
INVIOLABILITY	1044.809561	INFORMAL NOTICE	988.064314
JUDICIAL CONTROL	1044.597317	REDUNDANCY PAYMENT	986.998825
INDENTURE	1042.884508	PHYSICAL INJURY	986.902285
JUDICIAL FUNCTION	1042.629353	ABSENT WITNESS	985.688404
STAND	1042.575026	NUISANCE	985.60759
ADDRESS	1041.740062	ABEYANCE	984.325454
INTENT	1041.635034	JUDICIAL DETERMINATION	977.384085
SUSTAIN	1040.113085	DOMINANT TENEMENT	974.915101
DULY	1039.427984		
MISFORTUNE	1037.253448		
WARNING	1033.613167		

DECEIVE	974.003605	EXEMPTION CERTIFICATE	925.817027
NORM	973.968418	PRIMA FACIE CASE	925.614147
DISSENTING JUDGMENT	972.075778	LAW REMEDY	925.033424
NEGOTIATE	971.706186	HMRC	923.708223
IMPROPER	969.928983	DELIBERATELY	923.322295
MISREPRESENTATION	968.017197	WIDE DISCRETION	922.845729
GENERAL PRINCIPLE	967.230411	VALUE JUDGMENT	921.467115
AGGRAVATE	966.836988	JUDICIAL EXEGESIS	921.215242
GAIN	965.221173	IRREGULAR	920.668997
UNBIASED JUDGMENT	964.465706	REPUTE	920.652442
ACTUAL AUTHORITY	963.12171	LEGAL ENTITY	920.150397
OBLIGE	962.740694	PERSONAL MITIGATION	919.680074
DEPOSITION	961.864612	CREDIBLE EVIDENCE	918.090749
REPRESENTEE	961.336331	WILLING	917.417918
ILLICIT	959.412297	ACCOMPLICE	916.982874
MISDEMEANOUR	958.461278	OVERPAY	916.544085
OPTION CLAUSE	957.393424	ARREST	915.86809
OFFEND	957.341678	DENOTE	914.347267
LAWFUL POSSESSION	955.579109	DEROGATORY	913.519294
COMPELLING	955.349334	DECLARATION OF INCOMPATIBILITY	912.813755
INQUIRE	954.719752	CRIMINAL INVESTIGATION	912.638859
UNCONSCIONABILITY	954.334128	BENEFICIAL OWNERSHIP	909.67677
STATUTORY AUTHORITY	952.699041	INDEMNITY BASIS	909.520251
UNBROKEN	951.969636	DEPRIVATION OF LIBERTY	908.301274
DIFFER	947.90802	AMPLE EVIDENCE	908.025834
POSSESSOR	947.078375	SEXUAL OFFENCE	907.141773
PHYSICAL DAMAGE	945.922868	INSTALMENT	905.699825
PROCEEDING	944.614403	ACCUSER	905.532867
EQUALITY CLAUSE	944.498474	PROCEDURAL OBLIGATION	905.105811
REFERABLE	943.735517	APPLICATION FOR LEAVE	903.433987
JUDICIAL SEPARATION	943.472143	COMPETENT AUTHORITY	902.30911
IMPOSSIBILITY	943.179141	IMPENDING	901.686712
JUVENILE	940.995026	INQUISITORIAL	901.070053
STATUTORY GUARANTOR	940.941837	ENGLISH JURISDICTION	900.646707
BATTERY CLAIM	939.60698	LEGAL FEE	900.137974
DISCIPLINARY PROCEEDING	936.795715	LEGITIMATE INTEREST	896.731915
DESIDERATUM	934.362257	MAIN WITNESS	896.46955
EXCEPTIONAL CASE	934.340997	RELEVANT FACTOR	895.956185
INTERCEPTION	933.166102	EXPOSE	895.589448
RELEVANT LEGISLATION	931.685343		
SURETIES	928.720759		
ANNOYANCE	927.39015		

STATUTORY PROTECTION	894.574082	EARLY JUDGMENT	840.440161
ALIBI WITNESS	894.107756	HOMOSEXUAL	840.163207
APPROVE	893.486099	COUNSEL	840.051609
UNFAIRLY	891.31534	SHAREHOLDER AGREEMENT	837.972567
SEIZE	891.176835	UNLAWFUL WAR	837.928096
WHIPLASH INJURY	890.467249	LIABILITY ORDER	837.17979
PERSISTENT	890.17214	CONTRACTING AUTHORITY	837.017683
LAWFUL MARKET	883.561915	RELIABLE EVIDENCE	837.011503
INVALIDATE	883.311678	INTERROGATION	835.905096
CRIMINAL CONVICTION	883.135707	LAWFUL SENTENCE	835.343087
INTERIM PAYMENT	881.058407	RULE	834.439606
EXPROPRIATION	880.073474	IMPEDE	834.16391
INCHOATE	879.695007	IMMIGRATION STATUS	831.73798
STEAL	879.556191	DECLARATORY RELIEF	829.336164
REVOCABLE	879.09819	UNLAWFUL DEDUCTION	824.119842
TERMINATION	873.914559	KNOWINGLY	823.233788
FICTITIOUS	872.438966	CRIMINAL PURPOSE	819.675208
MATERIAL CONSIDERATION	872.399185	INTERLOCUTORY INJUNCTION	819.539303
APPORTION	871.442628	PURPORT	819.422363
MALTREATMENT	870.926233	USURP	818.360878
DEDUCTIBLE	869.116545	LAW CLAUSE	817.862408
TERRITORIALITY	869.008783	BYLAW	817.123353
OUSTER	868.512868	OVERSIGHT	813.886282
PREMEDITATION	865.258367	UNENCUMBERED	813.301864
IMPROBABLE	862.881497	LEGAL PROCESS	810.518903
CONSTRUCTIVE TRUSTEE	859.547592	PURSUER	809.275044
DEFENDANT	859.514439	PECUNIARY ADVANTAGE	808.012749
LEGITIMATE PURPOSE	858.671532	LAWFUL OBJECT	807.401102
LEGAL LIABILITY	858.356483	MENTAL DISABILITY	806.257665
ALLOW	858.290257	DOCUMENTARY EVIDENCE	805.44603
INTRUSIVE	856.576902	NEGLIGENT ACT	804.588118
RECUSE	854.498045	SPECIAL RULE	804.35599
FRAUDULENT DEFAULT	854.300139	VEXATIOUS LITIGANT	803.713279
STATUTORY BASIS	848.931982	EXPRESS PROVISION	802.518049
ABSOLUTE RULE	847.372661	QUESTION OF FACT	802.212447
STATUTORY LIMITATION	846.899215	IMPUTATION SYSTEM	800.893913
BARE TRUSTEE	845.490542	OFFENSIVE WEAPON	800.551352
EHRR	843.608984	DISCONTINUE	800.502535
STATUTORY CRITERION	843.186236	INADVERTENT	799.843455
LEGITIMATE INTEREST	842.872073	BINDING AGREEMENT	799.275741
UNLIMITED	841.20167	ORIGINAL CLAIM	795.179729

AUTHORITY	795.115983	CONCEDE	765.385045
UNBIASED	794.039785	OVERRULE	764.383545
MISAPPROPRIATE	791.847273	MERIT	763.972761
PROBATIONARY	791.672673	LOCK	763.680799
OFFICE	790.569787	LEADING JUDGMENT	762.561443
APPELLATE TRIBUNAL	789.586673	EX FACIE	761.92568
OUTPUT TAX	789.567285	ARBITRATION PROCEEDING	761.718352
ACCRETION	789.138106	DERELICTION	761.581965
UNWARRANTED	789.084205	OMBUDSMAN	761.550048
SUICIDE RISK	786.632742	ESTOPPEL ARGUMENT	761.251429
FORENSIC EXAMINATION	786.526192	LAWFUL IMMIGRATION	760.900816
VARIANCE	784.981309	FRESH NOTICE	759.614543
MARRIAGE CERTIFICATE	784.90395	NOTICE OF ADJUDICATION	759.576213
IMMOVABLE	783.660866	CONSTITUENCY	758.77258
DESTITUTION	783.535734	JUDICIAL REMEDY	757.738482
EVADE	783.13556	LIABILITY PRINCIPLE	757.021219
MALICIOUS FALSEHOOD	782.294157	EXCLUSIONARY	756.402048
IMMINENT	782.009581	ILLEGITIMATE	755.882776
INHIBITION	781.490123	DISMISSAL	755.717514
THIRD PARTY	781.356866	GRATUITOUS	755.394332
BREAK CLAUSE	780.427507	JUDGMENT	754.385714
MARITAL COHABITATION	779.221423	POTENTIAL WITNESS	753.833383
LEGAL PERSON	778.611969	CORPORATE TRUSTEE	752.642484
ANTECEDENT BREACH	778.540714	RIGHT OF ABODE	752.36957
EMPLOYMENT CONTRACT	777.131	RELEVANT RULE	752.342125
TORTIOUS LIABILITY	774.984086	COMPELLING EVIDENCE	752.290534
JUDICIAL EXAMINATION	774.760507	JUDICIAL SCRUTINY	752.058728
CONTRADICT	774.137357	JUDICIAL RATE	750.526434
SCINTILLA	773.578962	ADEQUATE REMEDY	750.4575
UNCONDITIONAL	773.01304	CODIFY	748.999424
FUNCTUS	771.788724	JUDICIAL RESTRAINT	747.487518
MISINTERPRETATION	770.438396	DISPROPORTIONATE INTERFERENCE	743.915228
MOVEABLE PROPERTY	769.805247	IRREMEDIABLE	743.905781
DISABILITY PREMIUM	769.712883	UNREASONABLE CONDUCT	743.55545
PRATIQUE	768.2301	CIRCUMSTANTIAL EVIDENCE	743.103628
DRAFT JUDGMENT	768.148275	EMPLOYMENT RELATIONSHIP	743.042851
ANONYMITY ORDER	767.954843	TRUTHFUL WITNESS	742.97973
MATERIAL BREACH	767.280112	ORDER FOR COST	742.232552
VAT FRAUD	766.191531	INTERIM INTERDICT	741.895836
BANKRUPTCY PROCEEDING	765.446324	SERIOUS HARM	741.806206
		REASONABLE	740.883817

CONCLUSION	
SOIL	740.551989
ADMINISTRATION OF JUSTICE	739.066301
REAL EVIDENCE	738.369639
LAW DOCTRINE	738.290289
ENURE	737.644411
PARAMOUNT CONSIDERATION	737.081195
SUBSTANTIAL DAMAGE	736.811287
FURTHERANCE	736.165495
CURTILAGE	735.88927
CALUMNY	735.606742
UNSOUNDNESS	735.422058
SSA	734.121762
STATUTORY ENTITLEMENT	732.540175
PHYSICAL HARM	729.874573
AWAIT	729.465324
INTANGIBLE	729.350274
EXTENDED SENTENCE	728.730286
MISJOINDER	728.518885
AFFILIATION	725.519469
BINDING CONTRACT	725.49368
ACT OF DISCRIMINATION	725.124287
STATUTORY TORT	724.866013
UNADMINISTERED	724.487131
APPEAL PROCEDURE	724.056058
INJURY PROCEDURE	722.285825
JUDICIAL INQUIRY	721.93225
PENSIONER	721.635758
INCOMPETENCE	721.519955
LIQUID DEBT	721.439093
GIVE JUDGMENT	720.688461
STATUTORY JURISDICTION	720.597808
RELEVANT WITNESS	718.121441
LEGAL TEAM	717.587043
ALLEGED OFFENCE	717.342954
DETER	716.850034
PRUDENCE	716.327514
LEGAL ESTATE	715.18431
RESIDE	713.281421
CAUSATIVE CONDITION	711.497845
ADMINISTRATIVE ACT	710.429573

ABRIDGE	710.056864
ABSOLUTE RIGHT	709.408668
TAX EVASION	709.40079
SHERIFF	709.080259
COUNCIL TAX BENEFIT	707.751979
ILLEGAL ENTRY	707.410176
PECUNIARY INTEREST	706.564285
SECOND JUDGMENT	706.046208
PURPOSIVE	705.524696
UNDERLEASE	705.293221
ARRAIGNMENT	703.677454
ILLEGAL ACT	702.572082
VIOLATE	702.261345
DISFIGUREMENT	701.666903
DISCIPLINARY PROCESS	701.407164
LEGAL COST	701.271329
SERVIENT	700.059141
PERIL	699.645955
PAUCITY	699.30504
NET ASSET	697.320916
CAUSALLY	697.247789
APPROPRIATE JURISDICTION	696.490454
CONSENSUAL	696.103231
SUBSIDIARY	695.980146
CIVIL CASE	694.017992
FALSE IDENTITY	694.000445
PRONOUNCEMENT	693.587871
INDICIA	693.283604
RECIPROCAL	692.567047
TAX RELIEF	692.547517
MANIPULATION	690.815539
ROYALTY	689.765374
STANDARD CONDITION	689.300424
DELEGATION	687.179802
ADDITIONAL CONSIDERATION	686.728222
INTERIM INJUNCTION	685.88646
CLANDESTINE	685.048437
DETAILED EVIDENCE	682.034562
ULTIMATE QUESTION	680.298032
ANCILLARY LIQUIDATION	680.082807
SUBSTANTIATE	680.040515
CIRCUMVENT	679.587575

GIVE EFFECT	679.307717		DISAPPROVAL	653.58209
EVIDENTIAL BASIS	678.925679		PATENT	653.372127
HEIR	678.132826		CAPITALISATION	652.753993
FLAGRANT DENIAL	678.122699		IRRECOVERABILITY	652.409275
EMBARGO	677.606758		WARN	652.276948
JUDICIAL SYSTEM	677.472656		CONCEPTION	652.218137
APPEALABLE	677.310891		PROBATE DUTY	651.611985
LEGAL CONSEQUENCE	677.302725		PRELIMINARY PROOF	649.748131
UNDERVALUE	677.065888		CRIMINAL CULPABILITY	649.561335
SUSPECT	675.859814		UNMERITORIOUS	648.237567
UNDUE PRESSURE	675.248978		FULL JURISDICTION	646.773908
CULPABLE NEGLIGENCE	675.203465		CONFISCATION PROCEDURE	646.693843
INMATE	674.273521		PERSONAL SERVICE	646.551088
REVERT	674.017417		REASONED	646.362689
AUTHORITATIVE	673.863254		LEGAL ORDER	644.922619
RELATIONSHIP OF TRUST	673.148153		CRIMINAL SANCTION	644.708066
PROSECUTION WITNESS	671.695943		COUNTERPART	644.63464
LAWFUL ORIGIN	671.603954		RELAXATION	644.494236
RECOUP	670.731526		TENANCY	644.460004
CONCURRENT DUTY	670.295352		ORDER	643.619051
RUBRIC	669.152147		ABSCONDER	642.957822
EXPEL	667.637258		UNIDENTIFIED WITNESS	642.448672
DILUTION	667.175113		INTERLOCUTORY APPEAL	641.552649
ELIMINATE	666.497056		OBSCURE	641.456436
STATUTORY INQUIRY	666.373585		ATTORNEY	641.359375
MOLESTATION	665.553882		PRESUME	641.012149
INDUCE	664.992787		MATRILINEAL DESCENT	638.38994
MISAPPREHENSION	663.760548		PRACTICABILITY	637.712011
REINSURANCE CONTRACT	661.98407		ALLEGED DISCRIMINATOR	637.274874
INDEX OFFENCE	661.607657		PARTICULAR WITNESS	637.057846
ADDITIONAL LIABILITY	660.479338		UNIDENTIFIED WITNESS	636.918495
LORD JUSTICE	660.251264		JEOPARDISE	636.13123
DOWRY	659.983431		AUTHENTICITY	635.720333
ILLIQUID	659.443016		PTWD	634.818288
RELEVANT DECISION	658.90953		ACTUAL WORDING	633.553777
SATISFACTORY EVIDENCE	658.293509		SUBSTANTIAL RISK	633.33132
ORDINARY RULE	655.949464		SUBSTANTIALLY	633.22951
REVIVE	655.732964		OBSTRUCT	633.21992
CONTRACT OF SERVICE	655.437718		CONDEMN	632.827366
AVOIDANCE OF DOUBT	654.692315		SETTLEMENT	632.576504
CAROUSEL FRAUD	653.863382		SUBSTANTIVE BREACH	632.199386

APPELLATE AUTHORITY	631.77029	REBATE	609.833094
NOTORIOUS	631.27852	QUESTIONABLE	609.530754
ANTEDATE	630.239385	MISCONCEIVE	608.143538
SOLICITATION	630.088707	CONNIVANCE	607.93267
SEIZURE	629.61145	DUTY OF TRUST	607.691336
PUTATIVE EMPLOYEE	629.241758	PROSTITUTION	607.297072
NOMINAL RATE	628.405187	CURIAL	607.102459
STATUTORY NOTICE	627.632797	GOOD JUDGMENT	606.956737
INCONGRUITY	627.36706	PROSECUTOR	606.9563
LEGAL EMPLOYMENT	627.219921	ACTIONABLE BREACH	605.970583
CONTRACTUAL OBLIGATION	626.839042	DOMESTIC COURT	605.913835
EXCLUSIONARY RULE	625.318451	DESTITUTE	605.847956
PUNISHABLE	624.390067	ALLEGED INFRINGEMENT	605.066841
DELIBERATION	624.348155	RESTORE	604.781541
ACTUAL SEISIN	622.299887	PRECURSOR	604.771704
PROTECTION	622.232365	STATUTE BILL	603.833445
EVIDENTIARY	622.106349	ASSIGNOR	603.603213
LAXITY	621.112141	CONCURRENT SENTENCE	603.200126
SAR	620.773282	FULL JUDGMENT	602.623304
PECUNIARY LOSS	620.250054	FULL ARGUMENT	602.5018
LEGAL	619.79102	ADJUDGE	599.582358
PARTIAL IMPUTATION	619.381395	PAROLE	599.399399
ABSENCE OF EVIDENCE	618.956357	DOUBTFUL	598.849099
DISCLOSE	618.714845	REMEDIAL WORK	598.840797
SPECIFIC INTENT	618.578772	PROPRIETARY CLAIM	597.777524
TAXABLE PERSON	618.210823	LINKAGE	595.843997
SECONDMENT	617.476347	TERM CONTRACT	595.734941
DECEASED PERSON	617.400529	COVER	595.308654
LEGAL ISSUE	617.133914	WITNESS SUMMONS	595.046835
TAP	616.988583	SPECIFIC EVIDENCE	593.337363
CLARIFICATION	616.067398	COMPILATION	592.294194
UNDUE INTERFERENCE	615.712876	NUPTIAL CONTRACT	592.196911
UNREASONABLY	615.589418	WILFUL DISREGARD	591.925406
FAMILY HOME	615.06849	PERMANENT ACCOMMODATION	588.598953
PATENT INFRINGEMENT	614.853541	CIVIL CLAIM	587.884093
OBVIOUS ERROR	614.742916	WILFUL NEGLECT	586.295005
RENT LIABILITY	614.534975	INSPECT	585.997702
LEND	614.459666	INDICTABLE	585.098614
JUDICIALISATION	613.956007	APPARENTLY	583.571413
VERACITY	613.282225	WRONGFUL DISMISSAL	583.510506
LAW DUTY	612.815389	RESTRAIN	582.705724
TESTIFY	611.160104	QUALIFIED EXEMPTION	582.535362
FRESH INQUEST	609.888961	RELEVANT	582.379701

DOCUMENT	
LAWFUL TAX	582.145903
LEGAL NATURE	582.059809
SOUND	581.926156
PREDECEASE	581.44267
DEMOTION ORDER	580.841103
ALLEGIANCE	580.73766
PRIMARY CASE	579.685691
SWEAR	579.328489
JUDICIAL CONSIDERATION	578.907212
EX P.	578.167618
DILATORY	577.169193
TRIAL DATE	576.445126
CRUX	576.325919
STATUTORY EXPRESSION	575.40516
EXTINGUISH	574.90924
UNOPPOSED	574.736132
REFERENDUM	574.667969
ADMINISTER	574.294258
ADVERSE INTEREST	573.760721
LAWFUL LITIGATION	573.224824
LEGAL ERROR	572.810087
LEGAL PERSONALITY	572.439888
UNLAWFUL TAX	572.234945
FABRICATION	572.210741
OUTLAY	571.770862
PARENTAL RIGHT	571.590204
PREJUDICE	571.173473
FALSE DOCUMENT	570.376284
CRIMINAL PROCESS	570.057129
LAWFUL USE	569.764447
TRUTHFUL	568.514727
UNANIMOUS JUDGMENT	568.272651
RATEPAYER	567.748278
STATUTORY CAP	567.72736
RESUME	566.797418
DIVULGE	566.388519
PROPERTY DAMAGE	566.23864
BANKRUPTCY PETITION	565.491384
SOLE TRUSTEE	565.363777
CONSISTENTLY	565.296146
BASIC RULE	564.96129
DELINQUENCY	564.526421

SUPERVISE	564.313643
PRESUMED	563.908968
SUSPICIOUS	562.29283
RECORD	562.2324
CONVERT	561.690079
VEIL	561.678174
LAWFUL DIVIDEND	560.775561
INCLINATION	560.521824
DISCOVER	559.446001
OVERSUBSCRIPTION	559.033197
GAOLER	558.556828
PERSECUTE	558.404525
RIGHT OF ACTION	558.208266
COMPETENT COURT	558.062351
LICENSE AGREEMENT	557.250339
COMMUNITY ORDER	557.11275
COMPATIBILITY	555.690441
FOUL	555.677745
LEGISLATIVE PROVISION	555.574116
TITHE	554.256054
OBSTACLE	553.999003
FORESIGHT	553.704195
UNSECURED CREDITOR	553.062649
SUBSTANTIAL SUM	552.925232
INTERPRET	552.638481
DEMURRAGE	551.957087
FURTHER CONSIDERATION	550.617045
LAWFUL TERMINATION	550.408473
CIVIL PARTNER	549.99248
DISHONESTLY	549.311972
IMMOVEABLE	549.234563
INCORPOREAL	548.827842
QUESTION OF LAW	548.190761
INADMISSIBLE HEARSAY	547.85458
PROPERTY LAW	547.29482
PRONOUNCE	547.222852
CONCLUSIVE EVIDENCE	545.506892
ORDINARY CASE	545.468006
ABUSE	545.321349
APPELLATE	544.93395
ADMINISTRATIVE DISCRETION	544.707158

UNKNOWN	542.85683	TERMINATION	
UKHL	542.692579	REGISTRAR	521.035549
JUDICIAL FACT	542.027387	NORMAL PROCEDURE	520.490008
INCITEMENT	541.352785	IMPUTABLE	520.461801
RECLAIM	541.314941	TRAFFICKING	520.366092
ENFORCEABLE JUDGMENT	539.435915	SEXUAL ASSAULT	520.327896
BELONGING	539.179972	FORCIBLE	520.288582
VOID	539.018804	PRECINCT	520.286064
REASONABLE CERTAINTY	538.763564	VALUABLE CONSIDERATION	520.276412
JUDICIAL OATH	536.411154	AGGRESSIVE BEHAVIOUR	520.269345
NOTIFICATION REQUIREMENT	536.348875	VEXATION	520.082416
MORAL TURPITUDE	535.749219	JUDICIAL ROLE	519.266391
PROSECUTE	533.987445	ALIBI	519.005973
MATTER OF LAW	532.637024	MATRIMONIAL ASSET	518.082792
FRUSTRATE	532.503117	CAUSATIVE EFFECT	517.943887
LEGAL SYSTEM	532.236101	DEFERMENT	517.45161
ADULTERY	532.08194	LIMITED	517.295299
MANDAMUS	530.742498	ORDINARY COURT	517.253453
PRIMARY FACT	529.811088	UNREGISTERED	517.205129
EXCEPTIO	529.578887	LAWFUL PROPERTY	516.283103
EVALUATIVE JUDGMENT	529.560381	AMENABILITY	516.269907
PERTINENT	529.485248	QUOTATION	516.246381
INURE	528.732868	FORMAL CONTRACT	516.161133
TERMS OF CLAUSE	528.249212	PLENARY	515.723956
CONTRACTUAL DUTY	527.887518	DISABILITY	515.144797
PROVABLE	527.826626	BASTARD	515.056302
ENCUMBER	527.4671	CONFISCATION PROCEEDING	512.491643
BASIS OF PLEA	527.315862	NOTICE OF READINESS	512.16776
GRATUITOUS BENEFIT	525.782413	SHORT JUDGMENT	511.818746
ABSOLUTE DISCRETION	525.633948	MAIN ISSUE	510.885107
RESIGN	525.377396	ERRONEOUS BASIS	510.81044
DECISIVE RULE	524.626623	FINAL SENTENCE	510.653201
ANCILLARY LIQUIDATOR	524.603383	ABSOLUTE PRIVILEGE	510.315864
PERSONALLY	522.678856	MANIFEST ERROR	510.157564
LEGAL BURDEN	522.658446	STATUTORY COMPENSATION	509.987337
UNUSUAL CIRCUMSTANCE	522.451527	DO JUSTICE	509.205031
INTERNATIONAL COMITY	522.431157	PRIMARY LEGISLATION	509.127168
PROBATIONARY PERIOD	522.280337	ROTATION	508.84637
ADMINISTRATIVE ERROR	521.734445	CONTAINMENT	508.617704
		STATE	508.445254
DATE OF	521.498633	INCORPOREAL PROPERTY	508.359718

ENGLISH COURT	506.98468	DISRUPT	490.378496
ARBITRABILITY	506.859183	REPEL	490.280654
EXCISION	506.544044	EVIDENCE IN SUPPORT	490.108415
SALE CONTRACT	505.744002	REGULATION	489.75488
UNLIQUIDATED	505.538379	LAWFUL RIGHT	488.938075
BARE	504.961227	MATTER OF FACT	487.093385
AMPLIFICATION	504.457484	FRAUDULENT MISREPRESENTATION	486.721824
ONUS OF PROOF	504.398633	AGGRAVATING FEATURE	486.578146
PERSONAL OBLIGATION	503.38327	CIVIL ACTION	486.14359
INAPPLICABILITY	503.20295	INVESTIGATIVE	485.990109
DECISIVE EVIDENCE	503.13995	CORROBORATIVE EVIDENCE	485.232479
COMMISSIONER	503.100815	LEGISLATIVE COMPETENCE	485.026136
PRELIMINARY HEARING	501.232469	REPAYABLE	484.731826
SECURITY BENEFIT	501.225785	BREACH OF COVENANT	484.533492
UNREPRESENTED	500.582266	DISMISSAL CASE	483.537193
ELV	500.349542	UNLAWFUL CASE	482.488433
BLAME	499.944601	DEDUCE	481.750687
SPECIFIED OFFENCE	499.522372	REGULARITY	481.46903
LEGAL RELATIONSHIP	499.3247	THEREON	481.381811
NON SEQUITUR	498.113315	PENSION BENEFIT	481.351098
INCEST	498.0199	EXCLUSION ZONE	480.26623
ACTUAL FACT	497.996417	SUBSEQUENT CASE	480.14457
PUBLIC BODY	497.693271	EXCLUSIVE RIGHT	480.009261
GRAVE	497.284155	APPARENT BIAS	479.09609
TERM OF IMPRISONMENT	497.279506	LAWFUL MARRIAGE	478.913375
BEHEST	496.394453	WRONGLY	478.41923
PROFESSIONAL JUDGMENT	496.111411	UNLAWFUL DISABILITY	478.182861
ACCESSORY LIABILITY	495.591336	COMMISSION	478.152819
STATUTORY RULE	494.126199	OPEN EVIDENCE	477.490543
JUDICIAL STATUS	493.813608	RELIEVE	476.622323
JUSTICIABLE	493.800962	DENOMINATION	476.559207
ADDENDUM	493.453595	REGULARISE	476.350053
STATEMENT OF REASON	493.384844	DEPONENT	476.128025
PERSISTENCE	493.253422	CONFLICT OF INTEREST	475.966698
LAWFUL ACTIVITY	493.044755	EX TEMPORE	475.78641
DESCENDANT	492.984686	ARGUABLE DEFENCE	475.084192
ENUNCIATE	492.134291	LATENT DEFECT	475.062896
EXPIRE	492.070244	JUDICIAL OFFICER	474.462629
NEGLIGIBLE	491.551064	INDEMNIFICATION	473.849219
FOREIGN JURISDICTION	491.17162	PARAMOUNTCY	473.169824
SODOMY	490.722476	CIVIL COURT	472.939253

CPR	472.140716	ASSURE	459.419815
JUDGMENT PROCEDURE	471.759743	LOCUS STANDI	459.302476
INHERENT POWER	471.321328	MALPRACTICE	459.132296
EXEGESIS	471.192694	POSSESSORY	458.743296
INFEFTMENT	471.163102	LEGISLATIVE SCHEME	458.379007
AWARD	470.975049	EXPERT WITNESS	458.211914
REMEDIAL ACTION	470.615297	COMPLETE AGREEMENT	457.613139
INAPPROPRIATE BEHAVIOUR	470.041567	EXERCISE OF JUDGMENT	457.227869
CUSTOMARY	469.532485	INSOLVENT LIQUIDATION	456.811212
PRIMARY LIABILITY	469.131543	ILLIQUID CLAIM	456.557383
TENDERER	468.792305	JUDGMENT REGULATION	456.081173
LAW JURISDICTION	468.364989	DIRECTIVE	455.784687
SECTION	467.855494	RETROCESSION	455.63227
CONSTRAINT	467.67313	ABSOLVE	455.557907
DESERTION	467.472391	JUDICIAL PRACTICE	454.944658
TRANSSEXUAL	467.304305	DEPENDENCY	454.879054
PRIORITY DATE	466.863683	DEED OF TRUST	454.712829
EXCLUSIVE CLAUSE	466.854715	LBO	454.378073
BEARER	466.681999	MANDATORY SENTENCE	454.229754
INVIOLABLE	466.656973	PROFITABILITY	454.126541
COVENANTEE	465.900757	TYPESCRIPT	454.10567
UNFETTERED DISCRETION	465.663292	ASSIMILATE	453.867073
VALID NOTICE	465.574108	DEMARCATION	453.801671
MENACE	465.18495	SETTLE	453.210912
DUTY OF CARE	464.79894	INFANCY	453.116459
EWHC	464.032669	RESPOND	452.783352
SUBSTANTIVE RIGHT	463.768127	EXEMPLARY DAMAGE	452.088476
LEASEHOLDER	462.909054	RULE OF COURT	451.886186
NIL LIABILITY	462.218545	ANONYMOUS EVIDENCE	451.837888
TOLERANCE	462.144257	CONCURRENT JURISDICTION	451.63438
ACCRUAL BASIS	462.059661	DISCUSS	451.421273
ANCILLARY MATTER	461.799964	UNAUTHORIZED	451.414685
EXTRANEOUS	461.631602	REPRESENTOR	450.379186
PRELIMINARY PLEA	461.462135	KIDNAP	450.222022
EVASION OF VAT	461.432177	EXTINCTION	449.760639
RATIFICATION	461.321827	CERTIFIED QUESTION	449.67844
CONTRIBUTORY FAULT	461.301459	RELINQUISH	449.232971
SUBPARAS	461.235892	BAILIFF	448.652822
PATENT PROTECTION	460.438664	PARI PASSU	448.300838
LTD.	460.350744	LEGAL SEPARATION	448.204404
SUBSTANTIVE OBLIGATION	460.099355	LEGISLATIVE PURPOSE	447.846502
DISTRIBUTABLE	459.6839		
WRONGFUL CONDUCT	459.502487		

PHYSICAL RESTRAINT	447.794077	VALIDLY	435.571294
CONDITIONAL RELEASE	447.238706	ASYLUM APPLICATION	435.505711
INTELLIGIBLE	447.097988	EXTREMIST	435.499446
LEGITIMATELY	446.996788	CONJECTURAL	435.369627
SUMMON	446.308752	ULTERIOR MOTIVE	435.271967
COGNIZABLE	445.556436	COMPARABLE EVIDENCE	435.245122
UNFETTERED RIGHT	445.549376	CRIMINAL INTENT	434.642484
BLAMEWORTHY	445.382893	LAW RULE	434.572414
YOUNG PERSON	444.846443	LAST RESORT	432.927427
LEGAL REMEDY	444.762744	DOMESTIC LEGISLATION	432.799923
REVIEWABLE	444.37598	RECITE	432.609174
WILFUL DEFAULT	444.301958	UNCONSCIOUS	432.5502
GUARANTEE LIABILITY	444.288391	STATUTORY FORM	432.499954
DEFERRAL	444.2463	MATERIAL MISDIRECTION	432.421888
UNDISPUTED EVIDENCE	443.722616	MORATORIUM	431.542067
INTERNAL REVIEW	443.456182	STATUTORY RAPE	431.206866
REMARRIAGE	442.964613	EPG LISTING	430.920225
EVIDENTIAL BURDEN	442.957519	FACTUAL FINDING	430.820025
STATUTE BOOK	442.345245	JUDICIAL EVICTION	430.721723
TRIAL COUNSEL	442.215912	BROCARD	430.66158
TORTIOUS DUTY	442.199774	PROPRIETARY RIGHT	430.632777
CORRECT PROCEDURE	441.994492	ADMINISTRATIVE	428.896749
OBSERVE	441.793627	PRESCRIPTIVE PERIOD	428.268435
ZONING	441.553229	COMPLICIT	428.075683
UNSIGNED	441.330616	CHARGE	427.094244
FALSE ALIBI	441.040578	APPREHEND	427.059865
FORENSIC PREJUDICE	440.577767	AMPLIFY	427.014291
PUNISH	440.213006	DIRECT EFFECT	426.871296
STATUTORY REDUNDANCY	440.050656	PLAINT	426.80406
AVOIDABLE	439.92893	EXCULPATORY MATERIAL	426.748796
SUBSCRIBER	439.258821	KIDNAPPING	426.147746
PATENT AGENT	439.098189	POSITIVE EVIDENCE	425.312588
WAR CRIME	439.095494	DISPENSATION	425.111807
RULE OF LAW	438.731872	IMAGE	424.826463
VALID PATENT	438.505932	SOUNDNESS	423.448424
PRETEND	437.379721	PICKET	423.312179
CULPABLE DELAY	437.124246	USUFRUCT	422.269588
LAWFUL MANNER	436.585109	SOLICITOR	422.2688
FREE EXPRESSION	436.561318	ORAL JUDGMENT	421.793401
JUDICIAL ADJUDICATION	436.560134	SUBSTANTIVE APPEAL	420.317563
PRINCIPLED	435.917892	INFLICTOR	419.507705
LINE OF AUTHORITY	435.910684	LEGAL PROCEDURE	418.867431
INIQUITY	435.668088	PORTFOLIO	416.969692

ACCOUNTABLE	416.677186	ARBITRAL PROCEEDING	402.258323
JUDGMENT CALL	415.769861	DENUNCIATION	401.923714
MARITAL PARTNERSHIP	415.526588	CONTRACTUAL RIGHT	401.396278
ABET	415.162811	IRRETRIEVABLY	400.884203
PARENTAL RESPONSIBILITY	414.852113	JUDICIAL ORDER	400.869007
FRAUDULENT SCHEME	413.058505	RIPARIAN	400.784596
SPECIALIST TRIBUNAL	412.709846	INTERMITTENT CUSTODY	400.107626
PREDOMINANT PURPOSE	412.5995	PROPRIETARY ESTOPPEL CLAIM	399.725671
DEFICIENT	412.30825	INTERIM RELIEF	399.366989
LIMITED JURISDICTION	411.731259	CESSER	398.150215
CONTRIVANCE	410.694646	COURT ROOM	397.628882
FREEZING ORDER	410.20781	FOREIGN COURT	397.54036
DISPROVE	409.944206	POTENTIAL INFRINGEMENT	397.317394
DIXIT	409.689853	SUBSTANTIVE LAW	397.155761
FORBID	409.529813	ENROL	396.750601
UNDISCHARGED	409.473945	VIOLATION OF ARTICLE	396.245778
MISADVENTURE	409.439803	ABSOLUTE EXEMPTION	395.756533
ULTERIOR PURPOSE	408.772421	IMMIGRATION AUTHORITY	395.365909
ECJ	408.694963	IMMUNE	395.286192
END ALLOWANCE	408.652752	UNLAWFUL DIVIDEND	394.467778
RENVOI	408.522087	CIVIL COURT	393.872413
PERMANENT INJUNCTION	408.520026	FRAUDULENT CONDUCT	393.301603
TENANT	408.111902	SECRET	393.273189
BAILII	407.793289	REFERRAL	392.536382
LEGAL REQUIREMENT	407.647721	QUID	392.342063
EXPRESS DECLARATION	407.38308	CONTRACTUAL LIABILITY	391.784105
DEFRAUD	407.082749	BACKDATE	391.688454
LAWFUL DECISION	406.461496	UNFIT DEFENDANT	391.672236
PROCEDURAL IRREGULARITY	406.073536	ORAL TESTIMONY	391.653999
DISCREETLY	404.537015	IMBURSEMENT	391.156308
CONDEMNATION PROCEEDING	404.416462	DEPOSITOR	391.096966
REASONABLE FORCE	403.96138	PROCEDURAL JUSTICE	390.837148
BARE TRUST	403.679416	PROCEDURAL PROTECTION	390.665563
REPOSSESSION	403.31747	UNDUE HARDSHIP	389.343021
IRRECONCILABLE CONFLICT	403.160367	LIEU OF NOTICE	388.875716
ENFORCEABLE CONTRACT	402.631084	ANCILLARY DUTY	387.958375
JUDICIAL SUPERVISION	402.604519	EXAMINATION OF WITNESS	387.816415
SUMMARY PROCEEDING	402.425939	RELIGIOUS DISCRIMINATION	387.730202
		UNLAWFUL SEX	387.241483

CERTIFICATE OF INADEQUACY	387.205798	CAUSATIVE LINK	376.542831
FUNDAMENTAL PRINCIPLE	387.134674	SECONDARY LIABILITY	376.077475
RETIRE	386.715781	LEDGER	376.001782
NAME	386.606448	SUE	375.51088
UNLIKELIHOOD	386.469104	NEGLIGENT MISREPRESENTATION	375.442552
PUNITIVE ELEMENT	386.32775	MANDATORY ORDER	375.1864
BLAMEWORTHINESS	385.885897	USURPATION	374.651244
CIVIL	385.878489	PERSECUTORY	374.553757
NEGLIGENT CONDUCT	385.635253	DEED	374.148444
PROCEDURAL FAIRNESS	385.593791	LITIGANT	373.631258
SIGNIFICANT HARM	385.33509	PERSUASIVE AUTHORITY	373.434082
ANNUL	384.579199	JUDICIALLY	373.326115
WILFULLY	384.015187	DISOBEDIENCE	372.472737
QUORUM	383.579601	DELIBERATE ACT	372.18066
BYSTANDER	383.45472	PREPARATORY HEARING	371.113954
RESEMBLANCE	382.886827	CRIMINAL STATUTE	370.499708
CONFLICTING EVIDENCE	382.319484	TOKEN	370.452338
DISCIPLINARY OFFENCE	382.223541	LODGMENT	370.406654
PERSUASIVE EVIDENCE	381.460523	ABOLISH	369.94349
INDICTABLE OFFENCE	381.308887	UNAUTHORISED ENTRY	369.729551
FRINGE	381.170037	SICK LEAVE	369.708964
PERSONAL RESPONSIBILITY	381.067756	UNSUBSTANTIAL	369.200436
INJURY CLAIM	380.729072	INTERLOCUTOR	366.911628
OFFEROR	380.406642	DEMUR	366.816345
ADVERSARIAL	380.284593	POLICE WITNESS	366.526266
GIFT TRUST	379.921881	ARGUABLE POINT	366.026333
UNEXCEPTIONAL	379.883018	ACQUIESCE	365.636202
ANNOUNCEMENT	379.498647	POTENTIAL CLAIM	365.631045
JURISDICTION ISSUE	379.496878	INSURE	365.199727
SECURITY OF TENURE	378.739474	CONTRACT OF INSURANCE	365.033923
INJURIOUS	378.646362	APPELLATE PROCESS	364.931333
SEPARATION AGREEMENT	378.644963	PSYCHIATRIC EVIDENCE	364.575685
PROCURATOR	378.377677	DISQUALIFICATION ORDER	364.337887
PARTITION	377.747809	INTERLOCUTORY JUDGMENT	363.560874
LIFERENTER	377.741147	ESCROW AGENT	363.475663
PROFESSIONAL MISCONDUCT	377.708027	INDEPENDENT CONTRACTOR	363.268841
INEFFICIENCY	377.445105	LAW PRINCIPLE	362.884476
PREREQUISITE	377.346156	LICENCE PERIOD	362.168696
DOMESTIC CASE	377.316086	UNDUE BURDEN	361.916924
DETENTION ORDER	377.229488	ORDINARY CRIME	361.3736

EFFECTIVE CAUSE	361.257342	IRLR	347.842955
RENTAL INCOME	361.167446	CONTRIBUTOR	346.808953
ENLARGE	360.650697	CONVINCING EVIDENCE	346.437753
MOVABLE PROPERTY	360.268648	LEGAL ANALYSIS	346.282313
UNAPPEALED	359.921867	CONSEQUENTIAL LOSS	346.167627
PHYSICAL LOSS	359.849188	ENSUE	346.105746
PHYSICAL DISABILITY	359.333312	RESIDUARY BENEFICIARY	346.015604
ALLEGED OFFENDER	358.783371	RELEVANT MATTER	345.72656
INFORMED	358.33483	ORDINARY PRINCIPLE	345.459028
PROMISE	358.095938	SUBSTANTIVE DUTY	345.109924
FORECLOSURE	357.735015	HEAD LESSEE	345.028462
PUBLIC TRUST	357.699681	FALSE WITNESS	344.534561
DUE NOTICE	357.386716	LEGAL CONSULTATION	343.276943
TAXABLE SUPPLY	357.227	FOUNDED	343.264814
ACTUAL NOTICE	357.205282	CONTEMPORANEOUS EVIDENCE	343.084778
LAWFUL TRADE	357.16191		
PERSONAL CLAIM	357.117347		
UNEQUIVOCAL	356.996369		
APPLICATION	356.859762		
SUFFICIENT BASIS	354.711476		
VESTED RIGHT	354.43649		
UNDERWRITER	353.289483		
TERRORIST ACT	352.204477		
PHYSICAL ABUSE	352.102191		
VALID CLAIM	351.992973		
BIGAMY	351.6494		
SPECIFIC PROVISION	351.611096		
VALUATION EVIDENCE	351.491905		
LAWFUL CLAIM	351.290367		
DEFICIT	351.288571		
SAME PARTY	351.151855		
CRIMINAL PENALTY	351.000205		
CONTRACTUAL JURISDICTION	350.736065		
GOOD PRACTICE	350.546894		
ACCEPTABILITY	350.254438		
INTEREST RULE	350.146202		
UNDISPUTED	350.121202		
LAW CLAIM	350.011672		
MONOPOLY	349.508258		
LIMITED LIABILITY	349.08303		
DISQUALIFICATION	349.031971		
JUDGMENT CREDITOR	348.9601		
MARGIN OF APPRECIATION	348.795871		
TEMPLATE	348.522989		

3.2.6. Applications of the term lists obtained

3.2.6.1. Some pedagogical considerations on the use of corpora in the ESP classroom

The potential applications of specialised vocabulary inventories are manifold. They can be employed by linguists, translators or ESP instructors as reliable sources of information for linguistic analysis, translation or language teaching. In this section, we will present different ways of exploiting our corpus and the term lists obtained from it after implementing the methods tested above within the field of English for Specialised Purposes (ESP) teaching.

The role played by specialised corpora in ESP is discussed by scholars like McEnery and Wilson who underline the fact that they meet the needs of ESP students better than general corpora "including quantitative accounts of vocabulary and usage which address the specific needs of students in a particular domain more directly than those taken from more general language corpora" (McEnery and Wilson, 1996: 121). They continue to assert their advantages in exposing learners to genuine language samples and acting as reference for scholars to review existing didactic materials. Schmitt (2002) affirms that their use might be beneficial regarding them as a valuable teaching resource as well as a useful tool to assess vocabulary acquisition. In addition, Gilquin and Granger (2010) insist on the importance of ESL learners' exposure to authentic materials based on corpora which also offer "a large number of authentic instances of a particular linguistic item" (Guilquin and Granger, 2010: 359) thus helping to identify their meanings depending on the context where they occur.

Conversely, Flowerdew (2009) criticises data-driven methodology due to its tendency to resort to decontextualised concordance lines extracted from corpora. This author agrees with Swales (1990) and Kaltembök and Mehlmauer-Larcher (2005) that "truncated concordance lines are examined atomistically" (Flowerdew, 2009: 395).

Tim Johns' (1986, 1991) work in this area is fundamental as he coins the term *data-driven learning* (DDL) emphasising the use of concordance lines extracted from corpora in the English classroom where students infer the rules of language by directly observing them. They are expected to "develop strategies for discovery –strategies through which he or she can learn how to learn–" (Johns, 1991: 1). Johns believes that, by discovering the rules of language underlying real samples extracted from corpora students become "language detectives" (Johns, 1997: 101). Hunston underlines the motivating character of this learning method which may help learners to remember already acquired patterns and also bring to the foreground "previously unnoticed patterns ... that a teacher [may have] overlooked" (Hunston, 2007: 170). Following Johns' first steps into DDL methodology, Sinclair (1991, 2003) continues to develop it further having been used as reference for over twenty-five years, and being considered as one of the most influential scholars in the area (McEnery & Xiao, 2011).

Though, in a way, highly motivating, Hunston (2007) also points at the disadvantages of employing "raw corpus" samples as the base for this type of methodology, that is to say, "the student and the tutor will look at the corpus together, without either of them necessarily knowing what they will find" (Hunston, 2007: 171). This author believes that, apart from being difficult to monitor as regards timing, there is little control on the part of the language instructor over the possible outcome of the activity. Hunston presents an alternative to this uncontrolled practice by referring to the design of corpus-based materials which may include selected examples to foster the acquisition of a particular grammar point bearing in mind the learners' proficiency level. Removing undesired examples that may result confusing (especially at earlier stages in language learning) is an excellent option to employ DDL methodology in a more effective way.

The term *DDL* is revisited by Boulton (2011), amongst other authors, who attempts to embrace all the different senses and uses of a concept whose definition by Johns was too wide to be systematised. He highlights the advantages of using this methodology which is capable of "empowering learners to explore language corpora and come to their own conclusions" (Boulton, 2011: 563).

In fact, the use of corpora is considerably widespread in ESP and ESL (English as a second language) teaching owing to the fact that they can contribute to a greater or lesser extent to second language acquisition yet, as far as legal English is concerned, the scarceness of didactic materials based on legal corpora is manifest. This is the reason why counting on such term lists as the ones presented below can be an excellent aid for the ESP instructor to complement, for instance, already existing materials such as textbooks showing how specialised language is used in real professional contexts.

Authors like Harwood (2005) review pro and anti-textbook arguments based on Allwright's (1981) assumption that the process of acquisition of a language is multifaceted and too complex to be accomplished by textbooks alone. Harwood recommends the use of corpora to act as support for the EAP (English for academic purposes) class stating that they should "be used as a launch pad for classroom research into how the linguistic item in question is used by experts and students in the learners' local context" (Harwood, 2005: 158). Furthermore, nowadays, ESP and EAP textbooks such as the Cambridge International English collection are more and more often based on specialised corpora for syllabus design to try and bridge the methodological gap suggested by Harwood.

Römer (2008) also supports this idea by presenting the results of several studies aiming at evaluating the effectiveness of DDL in the second language classroom. "These studies demonstrate that corpora nicely complement existing reference books

and that they may provide information which a dictionary or grammar book may not provide" (Römer, 2008: 120). Along these lines, McEnery and Xiao's (2011) review on the use of corpora in language teaching and learning is probably one of the most comprehensive ones written to date. Based on Leech's (1997) work, they describe three main focuses as regards the convergence between CL (corpus linguistics) and ESL:

> "That convergence has three focuses, as noted by Leech: the indirect use of corpora in teaching (reference publishing, materials development, and language testing), the direct use of corpora in teaching (teaching about, teaching to exploit, and exploiting to teach) and further teaching-oriented corpus development (languages for specific purposes (LSP) corpora, first language (L1) developmental corpora and second language (L2) learner corpora)" (McEnery & Xiao, 2011: 364-5).

The following section presents some proposals based on the literature consulted to exploit the numerous possibilities offered by specialised corpora and the vocabulary inventories extracted from them.

3.2.6.2. Direct applications of the term lists provided and the specialised corpus

Before starting with an actual proposal of activities, we decided to consult three legal English textbooks: *Professional English in Use: Law* (Brown & Rice, 2007); *Introduction to International Legal English* (Krois-Linder & Firth, 2008) and *Absolute Legal English* (Callanan & Edwards, 2010) so as to decide on questions such as text coverage, the coincidences between our term inventory and the one extracted from the textbooks, or the relevance of our term lists in comparison with the one obtained from the latter. This was done to ensure the usefulness and representativeness of our term lists in comparison with the legal English textbooks used as reference which cover varied genres and legal areas other than law reports.

The first step consisted in scanning and processing the textbooks using an OCR software. Then, the texts obtained, which contained 196,245 tokens, were stored in raw text format and processed with *Wordsmith 5.0* (Scott, 2008) resulting into a type list of 14,686 items that could be analysed and compared with the ones based on *BLaRC*, our legal corpus (the set of texts obtained by scanning the three textbooks will be referred to as *LeG-TeXT* henceforth). We concentrated solely on SWTs to facilitate the comparison and the automatic search for concordance lines employing the *Concord* tool included in Scott's (2008) *Wordsmith's* package.

After extracting and validating the STWs in *LeG-TeXT,* having applied Drouin's (2003) ATR method, *TermoStat,* it was attested that 67% of the SWTs identified were already present in the term lists obtained from *BLaRC* using the same ATR method, a considerably high percentage taking into account the fact that the textbooks employed as reference deal with many different types of both private and public legal documents and topics apart from judicial decisions. Furthermore, the documents used in the textbooks examined are usually adapted to fit into CEFR[19] level B2 to C1, which makes such overlap percentage even more relevant as *BLaRC* is made up of authentic language samples.

LeG-TeXT was also analysed with Heatley and Nation's (1996) software *Range* adapting our term list (the SWTs obtained from *BLaRC*) to become a baseword list used as reference by the software (instead of employing the ones provided by default with the software programme from *GSL[20], AWL[21]* or *BNC*) with the purpose of establishing the percentage of running words in *LeG-TeXT* covered by it. Surprisingly, the specialised terms found in BLaRC covered 12.37% of the running words in the textbook corpus,

---

[19] Common European Framework of Reference for languages.
[20] Michael West's (1953) *General Service Vocabulary List.*
[21] Averyl Coxhead's (2000) *Academic Word List.*

nearly three times as much as the expected percentage of text coverage established by Nation for specialised vocabulary.

According to Nation and Waring (1997), knowing the most frequent 2,000 words included in West's (1953) *GSL* enables us to understand approximately 80% of the words in any text. Nation (2001) classifies vocabulary into four different categories depending on their level of specialisation: general words, which provide ca. 80% of text coverage (or *text range*, as Nation puts it); academic words, included in Coxhead's (2000) *AWL* which can cover around 10% of the words in any text; technical words, which cover approximately 5% of the tokens in the corpus; and low frequency words, that is, those which do not fit into any of these categories, which would cover the remaining 5% of words.

Nevertheless, the specialised terms in *BLaRC*, which would fit into Nation's category of technical words (5% predicted text range), cover almost three times as many words as it would be expected. Probably, the fact that legal terminology is often employed outside the legal domain can explain this finding. Actually, as shown in the introduction, after processing the lists of terms identified in *BLaRC* with *Range,* almost half of the specialised vocabulary in those lists was found in West's (1953) *GSL* and Coxhead's (2000) *AWL* as well as amongst the most frequent 2,000 words of *BNC,* which would possibly justify that a greater number of tokens in *LeG-TeXT* was covered by our term inventories. Such overlap could actually be explained by the fact that almost half of them are either shared by the specialised and general fields without changing their meaning or they acquire a new specialised meaning in the legal context.

In spite of the overall coincidences between the full term lists extracted from *LeG-TeXT* and *BLaRC,* once they were processed applying Drouin's (2003) *TermoStat* and owing to the differences between both text collections, as figure 12 indicates, the

level of specialisation of the 20 top terms in *LeG-TeXT* differs greatly from the same value in *BLaRC*. The highest ranking term in the former, *contract* (S[22]=101.39), ranks 78th in the latter (S=31.34), whereas a word like *court,* whose specialisation level reaches 61.48 in *BLaRC,* only displays 46.73 for the same parameter in *LeG-TeXT,* probably due to the relevance of the term *court* within judicial decisions, where this word is constantly employed as reference for case citations, amongst other uses. Conversely, a word like *contract,* highly relevant in *LeG-TeXT,* does not appear to be so outstanding in *BLaRC,* where contract law is just one of the many law branches which the cases heard at British courts belong to.

The most striking differences can be found amongst words like *lawyer* (*LeG-TeXT* S=57.56; *BLaRC* S=-7.76) or *client* (*LeG-TeXT* S=50; *BLaRC* S=3.8), which are identified as highly specialised terms in the textbook collection while they would have not been included in the term list obtained from *BLaRC* by *TermoStat* due to their low coefficient. Probably, their more general character (they can be employed in everyday English not requiring any specialised knowledge of the legal field) might justify this huge difference since textbooks are adapted to the learners' level and law reports are authentic texts which comprise highly technical vocabulary.

The graph below also reveals certain coincidences amongst words such as *liability* (*LeGTeXT* S=50.97; *BLaRC* S=30.92)*; breach* (*LeGTeXT* S=50; *BLaRC* S=36.04)*; damage* (*LeGTeXT* S=49.37; *BLaRC* S=26.88) or *party* (*LeGTeXT* S=48.57; *BLaRC* S=25.16) whose levels of specialisation in both corpora are similar possibly owing to their reference to common concepts related to judicial proceedings which may as well be part of the contents in a legal English syllabus focusing, for instance, on civil or criminal law processes.

---

[22] *SL* stands for *specialisation level*, according to Drouin's (2003) ATR method.

SPECIALISATION LEVEL
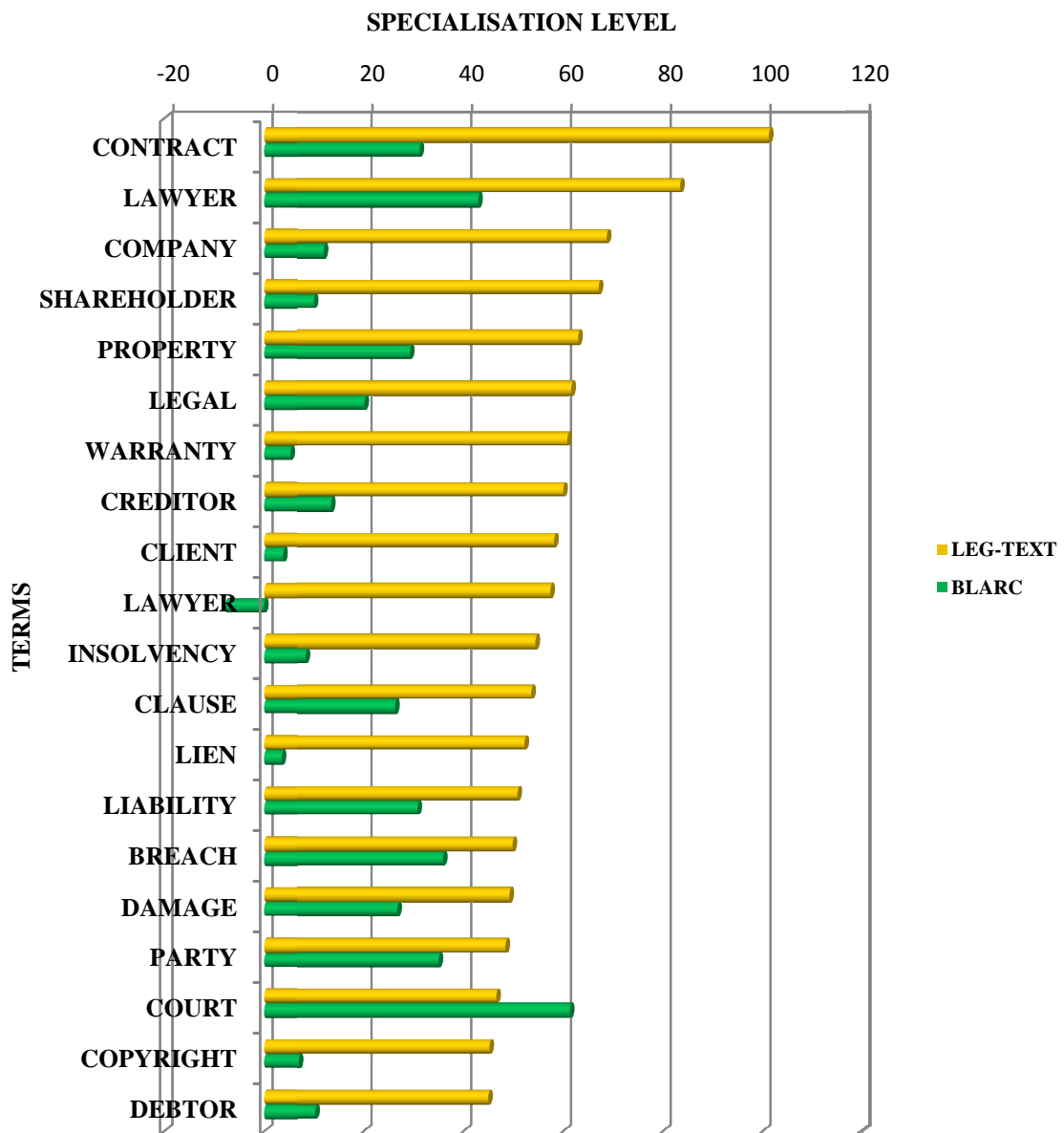
Figure 12. *Top 20 terms identified in LeG-TeXT using TermoStat and corresponding values in BLaRC*

Nonetheless, in spite of the differences amongst the words sampled above, the overall number of coinciding terms and the percentage of text coverage provided by the term inventories extracted from *BLaRC* might be indicative of their usefulness as

support material for the legal English class. Let us then suggest some activities that could be planned using our legal corpus as a source of information to develop them.

Numerous authors (Johns, 1991; Aston, 1997; Leech, 1997; Tribble & Jones, 1997; Brodine, 2001; Pérez & Cantos, 2004; Cotter, 2006; Fuertes-Oliveira, 2008; Rodgers, 2011, to name but a few) have carried out experiments using a DDL methodology to plan and evaluate different types of activities which focus on diverse language levels and learning skills, for instance:

o Studying the differences between synonyms like *convince* and *persuade* based on the analysis of the grammatical pattern found in the concordance lines provided (Johns, 1991).

o Finding the most frequent collocates of a technical word like *cancer* in a specialised corpus to try and establish a typology of the term (Rodgers, 2011), followed by oral and written discussions employing the terms found.

o Looking up words in a dictionary and contrasting their different senses with the information obtained from the concordances offered (Fuertes-Oliveira, 2008).

o Studying derivational processes by which words are formed stemming from a base or headword (Cotter, 2006).

o Evaluating term acquisition through the use of fill in the gaps exercises based on a corpus (Cotter, 2006).

o Promoting self-discovery by developing L2 learners' awareness of their own oral production (Pérez & Cantos, 2004).

As already stated, the activities suggested below could be employed as a complement to other existing teaching materials such as textbooks, thus, they are adapted to suit the competence level established for the three legal English textbooks

employed as reference which is intended to help the students progress from CEFR level B2 to C1, whereby they are expected to be able to understand specialised texts and communicate efficiently in a specialised environment.

For the accomplishment of the tasks proposed, students should be instructed in advance to use concordancers[23] so as to be able to easily access the information requested from the corpus used as support for the legal English class (*BLaRC* in this case). They should learn how to generate concordances, identify collocates, sort the concordance lines depending on their preferences, apply stop lists whenever it was required, adjust the settings for the identification of collocates, and so forth, so that the data provided by the corpus can be handled by them autonomously and exploited in as many ways as possible.

One of the activities that could be planned to develop the learner's awareness on the morphological structure of legal terms, which would also contribute to make them explicitly reflect on the processes underlying word formation, would be asking them to try and guess what terms would stem from a list of the most relevant ones found in their textbooks, that is, asking them to try and form part of their word families (Bauer & Nation, 1993). Before starting with the activity, it would be necessary to make certain morphological rules explicit as regards the use prefixes and suffixes putting special emphasis on typically legal ones such as *counter-, cross-, -ant,* etc. to facilitate and control the task.

Words like *appeal, decision, claim, law, jurisdiction, statute, liable, trial, act* or *crime* form other terms by derivation whose usage learners would have to attest through the search of concordance lines in *BLaRC*. These concordances would serve not only to confirm their guesses, but also to study their context of usage and meaning. The

---

[23] The concordancer employed in this case has been the *Concord* tool included in Scott's (2008) *Wordsmith 5.0.*

concordances below illustrate the use of some of the legal terms belonging to the word families of:

APPEAL:

- Permission to appeal against the quantum of damages was refused by the Court of Appeal but granted by this Court. The respondents were subsequently granted permission to <u>cross-appeal</u> against the finding that they were liable in trespass.

- The appellants have brought this appeal in order to challenge the finding of the majority of the Divisional Court that RIPA was intended to extend to legal or medical consultations. The respondents did not <u>cross-appeal</u> against the making of the declarations, although their counsel did attempt to argue that the surveillance was proportionate, claiming to be able to do so on the terms of the certificate.

CLAIM:

- I would reject the suggestion that a <u>counterclaim</u> against a public authority on the ground that it has acted (or proposes to act) in a way that is made unlawful under section 6(1) of the 1998 Act should be regarded as having been made under section 7(1)(b).

- The defence of opinion is lost where a <u>claimant</u> proves that the defendant did not act honestly in publishing the opinion complained of.

LAW:

- The underlying purpose of the Act, as I have already analysed it, reinforces that conclusion. "<u>Lawful</u>" in this context means having leave to enter.

- I further guarantee that I will not indulge myself into any <u>unlawful</u> or illegal activity in the United Kingdom."

TRIAL:

- These included admitting evidence of <u>pre-trial</u> statements made before a judicial authority and preserving the anonymity of witnesses.

- That order, so long as it stands, would prevent the BBC from broadcasting the circumstances of D's acquittal and discussing the possibilities of his future <u>retrial</u> save on an entirely anonymous basis.

ACT:

- The court said that it had also been accepted that, in view of the principle of legal certainty, a constitutional court may set a time-limit for the legislator to <u>enact</u> new legislation with the effect that an unconstitutional provision remains applicable for a transitional period.

- English law would not be determinative of the position under the law of the Cayman Islands after the <u>enactment</u> of the Cayman Islands Company Law 1989

On a syntactic level, the study of grammar patterns is another potential application of corpus-based activities within DDL. Learners could be asked to focus on the most frequent prepositions accompanying a set of legal terms by examining the concordances generated by them and concentrating on their collocate lists. They would be requested to study the main collocates of a group of words such as *appeal, claim, application, right* or *breach* with the aim of identifying those prepositions which the concordancer would present as their most relevant functional collocates. In order for the activity to accomplish its goal, only the immediate right collocates would be considered. After doing so, they would be offered different examples extracted from *BLaRC* to fill in the gaps with the appropriate prepositions to guarantee the validity of their observations.

The figures below illustrate the lists of the main collocates generated by *appeal* (figure 13)*, claim* (figure 14) and *breach* (figure 15)*.

Figure 13. *List of immediate right collocates attracted by appeal*

Figure 14. *List of immediate right collocates attracted by claim*

Figure 15. *List of immediate right collocates attracted by breach*

Focusing on a semantic level, it would also be interesting to study the contexts of usage of those terms which, according to numerous authors (as shown in chapter 4), might present certain difficulties for their acquisition due to their polysemic character. They are the so-called *sub-technical words*, defined and studied in greater detail in chapter 4, which characterise legal language and partly justify the great percentage of shared vocabulary between the legal and general fields. They may acquire a specialised sense when in contact with the legal environment causing confusion to ESP learners who might have already acquired them as part of their general vocabulary.

Learners would be given a list of these words taken from the corpus asking them to match the concordance lines obtained with the different senses of those words taken from a general and specialised monolingual dictionary: *The Oxford English Dictionary* (2002) and *Dahl's Law Dictionary* (Saint-Dahl, 1999). For this activity, they should also resort to a general corpus in order to identify the general meanings of the words given. There is a plethora of options, for instance, they could access other general

corpora either stored on their computers or offered online. Mark Davies' website (http://corpus.byu.edu) provides online access to varied general English corpora which could serve this purpose. The example below illustrates the most frequent senses of the *sub-technical* terms *party* and *offence* selected amongst the ones defined in the *OED* (2002) and the concordances obtained from *BLaRC,* our legal corpus, and *LACELL,* the general one:

- *PARTY*

MEANINGS (*OED*)	**CONCORDANCES**
1. A social gathering especially for pleasure or amusement	… if I were to plan a big party, or an anniversary or something like that, and I'd hope those would be jolly enjoyable days too …(*LACELL*)
2. An established political group organized to promote and support its principles and candidates for public office	I witnessed this over and over again, until I decided to join the Green Party in 1993 (*LACELL*)
3. A person or group involved in a legal proceeding as a litigant	Each party shall pay its own costs in respect of the issue of costs (*BLaRC*)

- *OFFENCE*

MEANINGS (*OED*)	**CONCORDANCES**
1. Annoyance, displeasure, or resentment	… I didn't take offence at the question, I think it was a perfectly fair question, … (*LACELL*)
2. A violation or breach of a law, custom, rule, a crime	The Director may withdraw or restrict access to the facilities in response to an offence or a suspected offence against these rules or to protect the services (*BLaRC*)

To finish with this section devoted to the didactic exploitation of corpus data, a subject-oriented activity is suggested. Using the MWT list generated by Nazar and Cabré's (2012) method, *Terminus 2.0,* we selected a number of these items that could help the learners understand and acquire, for example, such a concept as "types of

claim" in the UK. Then, they would have to clearly delimit the concept by differentiating the categories or types comprised by it. This could be achieved by providing concordance examples with gaps which they could fill in using the compound term list provided. Using a specialised dictionary would be recommendable as support to perform this task due to its greater complexity. In addition, they would probably have to consult the co-text of the concordances for a fuller understanding of the examples.

The tables below show the MWTs associated to *claim* and identified by *Terminus* (Nazar & Cabré, 2012). They have been arranged according to the level of specialisation calculated by the ATR method applied. Below, some of the concordances found in *BLaRC* are also offered:

CLAIM TYPES (MWTs)	SPECIALISATION (*Terminus*)
FRESH CLAIM	2767.60
POSSESSION CLAIM	1480.37
DERIVATIVE CLAIM	1426.45
UNFAIR DISMISSAL CLAIM	987.766
BATTERY CLAIM	939.60
PROPRIETARY CLAIM	597.777
CIVIL CLAIM	587.884
ILLIQUID CLAIM	456.557
RENEWAL CLAIM	414.608
EQUAL PAY CLAIM	407.413
PERSONAL INJURY CLAIM	380.729

CONCORDANCE LINES FROM *BLaRC*
… I don't believe it and therefore I am not going to regard it as a **fresh claim** …
… he could review the Council's decision to bring and maintain the **possession claim** on normal judicial review principles …
… there were special circumstances justifying the **derivative claim** which he seeks to bring …
… if he had been of the view that the **unfair dismissal claim** had not been made in time …
… prosecution in a civil court of the assault and **battery claim** would be "manifestly unfair" to him …
… her family was to ensure that the husband acquired no **proprietary claim** to shares in the wife's family …
… It is not necessarily an abuse to proceed with a **civil claim** in tort against a defendant who …

… that the rule that an **illiquid claim** cannot be pleaded by way of compensation to a liquid claim …
… the applicant's own evidence in his **renewal claim** form to benefit where the phrase 'emotional support' was used …
… in what circumstances, if any, can a claimant in an **equal pay claim** show that she is in the same employment as a man employed by the same employer at a different establishment in a different job?
… he had instructed a solicitor to lodge a **personal injury claim** against the company in respect of the injury to his hand …

An oral group discussion could follow this activity consisting in asking and answering questions about the meanings of the terms employed in it or the differences in this respect between the British and the Spanish legal systems. To conclude, an essay on this subject could also be proposed as a final task to complement their work on the concept of *claim* and *claim types* in UK law.

3.3. CONCLUSION

In this chapter, a comparison amongst ten different single and multi-word term recognition methods has been carried out. Such comparison was made using *UKSCC,* a 2.6 million-word pilot corpus of judicial decisions of the UK Supreme Court. Concerning precision (recall could not be calculated in all the cases because of the thresholds applied by the authors), the results differ greatly showing that Drouin's (2003) *TermoStat* and Nazar and Cabré's (2012) *Terminus* are the most efficient ones in identifying the terms in our legal corpus. They manage to recognise 73.45% and 71.50% true single and multi-word terms respectively on average. As a matter of fact, *TermoStat* reaches a peak of 88% precision for the top 200 CTs extracted whereas *Terminus* identifies 84.5% TTs within the same range. Such percentages were calculated by comparison with a gold standard, a 10,088 entry electronic legal glossary compiled from four different electronic sources and two specialised paper dictionaries, as stated above.

After selecting the most efficient methods tested on the pilot corpus, they were used to analyse *BLaRC,* an 8.85 million-word legal corpus of UK law reports compiled *ad hoc*. The lists produced were validated and the results varied slightly from the ones obtained in the analysis of *UKSCC*. In spite of using the same number of CTs as in the evaluation phase (2,000 CTs for SWTs 1,400 for MWTs), *Terminus* decreased considerably its efficiency by 7 points on average managing to identify 64.5% TTs and finding its peak at 78.5% for the top 200. As for *TermoStat,* the results were rather similar extracting 74.5% TTs and reaching a peak of precision of 87% for the top 200. Different options were also considered for evaluation, as shown above.

The validation of these different methods led to the production of two word lists, one of them showing the SWTs generated by *TermoStat* (table 3) and excluding those that *Terminus* also identified, and another one containing both the single and MWTs produced by *Terminus* (table 4). Needless to say that, in spite of the reliability of the gold standard used for evaluation, the lists were manually supervised to reduce noise to the minimum.

In the final section of this chapter, after reviewing different authors' views on DDL, some activities focusing on the legal lexicon were suggested to complement the ones offered in three different legal English textbooks using *BLaRC* as the source for such activities. They were focused on several linguistic levels, namely, morphological (concentrating on derivational processes for word formation), syntactic (grammatical patterns associated to certain legal terms), semantic (study of polysemic terms) and discoursive (oral discussions and essay writing suggested to reinforce the acquisition of compound terms). A final subject-oriented activity was also designed to try and focus on legal contents proposing a final oral discussion and writing activity to consolidate their acquisition.

To sum up, one of the advantages of resorting to ATR methods is that, due to the size of corpora nowadays, accessing such information as the specialised lexicon of a given language variety becomes a relatively easy task . Nonetheless, still much remains on the part of specialists to make the last decisions to discriminate terms from non-terms. When words have numerous senses, it is unavoidable to rely on the specialist's criterion to disambiguate them. As Lemay et al. point out, automatic methods might be of great help for terminologists to confirm their own intuitions and in particular, to "bring to their attention units that might have been considered as trivial and non-domain-specific" (Lemay et *al.*, 2005: 245).

The applications of the lists produced by these methods are diverse; they can be employed by translators, linguists or ESP practitioners. This chapter has also been intended to illustrate some of them.

CHAPTER 4
AN ANALYSIS OF SUB-TECHNICAL VOCABULARY

CHAPTER 4

4.1. INTRODUCTION

Having already generated a list of the specialised terms in *BLaRC* automatically, we will concentrate on a particular group of them which ATR methods do not differentiate from purely technical terms or even exclude from their lists. The character of these words is ambiguous and they pose considerable difficulties for their identification using statistical data owing to the fact that they are shared both by the general and specialised fields. Moreover, authors show their concern about them constituting a problem area in ESP teaching since they can be polysemic and often gain a new specialised sense in the technical context which may differ to a great extent from its general one.

Using Heatley and Nation's (1996) software *Range* to compare the list of single-word legal terms found in *BLaRC* with the most frequent words of English, it was found that that 40.47% of these terms are included amongst the most frequent 3,000 word families in West's (1953) *GSL* and Coxhead's (2000) *AWL*. The percentage is slightly higher, 45.41%, if compared with the *British National Corpus* thus confirming that almost half of the legal terminology identified in our corpus is shared with general English, as already affirmed in the introduction.

Let us then study this particular type of vocabulary commonly labelled as *sub-technical* by scholars.

4.2. SUB-TECHNICAL VOCABULARY: A REVIEW OF THE CONCEPT

As put forward by Sánchez (2000), the attempts to produce general vocabulary lists to be employed in language teaching can be traced back to early centuries, however, the

reliability of the earlier vocabulary inventories was questionable due to varied reasons. On the one hand, the sources they were obtained from, which clearly conditioned their representativeness, on the other hand, the criteria employed to organise them, or the polysemic character of some of the items comprised in them whose varied meanings were not accounted for. Sánchez (2000) underlines the relevance of the work by Thorndike and Lorge (1944) as the first ones to explicitly take into consideration the senses of the words in their lists with the aim of organising them.

Following Thorndike and Lorge (1944), West (1952) provides an inventory of the most frequent 2,000 word families in English: the *GSL*. "West's list incorporates important elements that had been the subject of discussion in the preceding years among 'basic vocabulary lists'. Particularly, a detailed specification of the senses of each word and the percentage of uses accounting for every one of the senses" (Sánchez, 2000: 8). According to Nation and Waring (1997) and Nation (2001), West's word families together with the 570 families from Coxhead's (2000) *AWL* cover 85-90% of the words in any text.

Nevertheless, there are words standing somewhere in between general and highly specialised vocabulary (which is almost exclusively employed in the scientific field) whose level of specialisation is hard to define, especially using quantitative criteria. According to Lan (2001), Cowan (1974) appears to have coined the term *sub-technical* to refer to context-independent words which are shared by different scientific disciplines. In his view, these vocabulary items must be specially emphasised by the EFL (English as a Foreign Language) instructor since they might cause certain problems in the teaching and acquisition of a second language due to their ambiguity. Cowan also introduces the concept *semi-technical,* which denotes something different. It refers to those lexical items which, in spite of belonging to general English, are

frequently used in technical texts. However, there is an earlier attempt to describe this type of words, as Lan remarks. Barber (1962) extends West's (1953) *GSL* "to bridge the gap between the basic *GSL* list and lists of strictly technical vocabulary items ... [Actually], this list … [provides] an inventory for explaining scientific ideas to the layman" (in Lan, 2001: 8), although Barber never mentions the term *sub-technical.*

Baker (1988) also labels this group of words as *sub-technical*: "The term *sub-technical* covers a whole range of items which are neither highly technical nor obviously general in the sense of being everyday life words (…)" (Baker, 1988: 91). As well as Cowan (1974), Baker addresses the question from a didactic point of view stating that this kind of words presents certain difficulties to the ESP instructor due to their obscure and unclear character. As a matter of fact, the author is not satisfied with the general/ technical division of vocabulary since there is plenty of evidence that many words in specialised fields belong neither in one category nor the other one. She offers an interesting classification based on varied authors' definitions of the concept *sub-technical.* Amongst the six categories established to classify sub-technical vocabulary, it is the last one, "items which are used in specialised texts to perform specific rhetorical functions" (Baker, 1988*:* 92), which appears to be more difficult to teach and acquire, according to the author. Students may be misled by a wrong interpretation of those linguistic elements which point at the writer's evaluation of the entire text or some relevant parts of it. Therefore, the ESP instructor will have to put greater emphasis on their teaching.

Similarly, Flowerdew shows his concern about the relevance of sub-technical words in ESP syllabus design. "They are words in general usage but which have a special meaning within the technical area" (Flowerdew, 2001:82). This is precisely why it is not the content teacher who will be in charge of working on their understanding and

acquisition but rather the ESP practitioner who must include them in their syllabus. Flowerdew refers to these words as *semi-* or *sub-technical,* as equivalent terms.

On the other hand, Chung and Nation offer a different perspective on the classification of technical vocabulary although they do not exactly employ the concept *sub-technical.* They rather provide a taxonomy including four different categories to distinguish those words which are purely general, that is, words "independent from the subject matter" (Chung & Nation, 2003: 105), from technical ones (those which are exclusively employed in the field of anatomy not being shared by the general field or other scientific disciplines). In it, they include categories two and three, namely, "words that have a meaning that is minimally related to the field of anatomy" and "words that have a meaning that is closely related to the field of anatomy. They refer to parts, structures or functions of the body, such as the regions of the body and systems of the body. Such words are also used in general language" (Chung & Nation, 2003: 105) which might certainly be identified with sub-technical words.

Wang and Nation (2004) go along these lines in their analysis of Coxhead's (2000) *AWL,* which is organised in word families around a single headword. They attempt to clearly distinguish the members of the same family from those words which are identical in form but utterly differ in their meaning: homographs. In order to do so, they produce a "semantic relatedness scale" (Wang & Nation, 2004: 291) by means of which they can distinguish whether the different semes of a word are related to each other in a way that it can be regarded as polysemic or if, on the contrary, they are completely unrelated thus being a clear instance of a homograph. They establish six semantic levels which the different senses of a word may fit into with respect to the base meaning. They go as follows:

"0 The meaning is the same as the base meaning.

1 The meaning is only slightly different from the base meaning.

2 The meaning is related to the base meaning with some changes.

3 The meaning is substantially different from but is still related to the base meaning.

4 The meaning is very distantly related and almost totally different from the base meaning.

5 There is no relationship at all between this meaning and the base meaning"

(Wang & Nation, 2004: 297).

Instead of concentrating solely on the semes of a given word, Trimble adds a quantitative perspective to the subject by defining sub-technical words as "those words that have one or more 'general' English meanings and which in technical contexts take on extended meanings" (Trimble, 1985: 129), showing high frequency levels amongst them. The author discriminates between those which are shared by all scientific fields without changing their meaning and those which activate a different one in a specific scientific field.

Finally, Farrell (1990) focuses words' frequency and distribution, concluding that, generally speaking, semi-technical words tend to be well distributed across disciplines also displaying high frequency counts. Conversely, technical vocabulary is not so well distributed although it should present high frequency levels in a specialised field.

All in all, authors tend to favour the use of the term *sub-technical* basically defined as shared vocabulary by both the general and the specialised fields or amongst scientific disciplines. Some of them also stress the relevance of the different senses of this type of words which acquire new meanings in technical areas. In addition, most of them underline their relevance in ESP instruction and the greater importance they must be given within the curriculum due to the fact that they might become an obstacle in the

learners' acquisition of the vocabulary in a any scientific field. Only Chung and Nation (2003) and Wang and Nation (2004) are more exhaustive as regards the delimitation of the semantic features of technical and sub-technical vocabulary in an attempt to analyse this lexical phenomenon from a different perspective, yet they do not label shared vocabulary as *semi-* or *sub-technical.*

Thus, taking all these different perspectives into consideration, and having observed a wide sample of sub-technical words (which will be referred to as such henceforth) taken from the legal corpus, *BLaRC*[25], they will be classified into three major groups related to their semantic features, frequency and fields of usage following Wang and Nation's (2004) proposal, namely:

1) Words denoting a legal concept which are frequently used both in the general and specialised fields not changing their meaning in the legal context: *judge, court, tribunal, law, prosecution, jury, legislation, robbery, theft, guilty, solicitor.*

2) Words often employed both in the general and specialised fields which change their meaning in the legal context sharing some semantic features with their original meaning: *charge, offence, sentence, claim, decision, grounds, complaint, dismiss, evidence, relief, record, trial, battery.*

3) Words occurring more frequently in the specialised field than in the general one which change their meaning in the legal environment acquiring a new meaning. Their new meaning is quite distant or completely unrelated to their general sense: *appeal, conviction, party, warrant, terms, act.*

This taxonomy excludes what Cowan (1974) labels as *semi-technical* vocabulary (differing from other authors' definitions of the same term), that is, general words which

---

[25] And also using LACELL as reference for their general meanings and usages.

are frequently employed in the legal field. The basic reason for this omission is that Cowan's idea of *semi-technicality* does not refer to words conveying any legal concept or acquiring a specialised meaning when in contact with the legal environment (the latter are considered *sub-technical*), although they are shared by both language varieties.

Finally, as already stated, ATR methods extract sub-technical vocabulary from specialised corpora assigning them a given weight within a list of CTs. No distinctions are made in this respect between highly specialised terms and this type of words since the context is not often taken into consideration. The fact that sub-technical words are shared both by the general and specialised fields makes them harder to spot by only focusing on statistical data, hence the need to examine their context of occurrence and usage so as to be capable of discriminating between their general and specialised senses.

This is the reason why such contexts will be explored in sections 4.3 and 4.4. applying, in the first place, Williams' (2001) lexical network model. Once the lexical networks of the words under examination are calculated, step 4 of the algorithm *Sub-Tech* (described in detail in section 4.3.2.) will be applied in order to place such words along a continuum of sub-technicality by comparing the networks obtained both from a specialised and a general corpus.

After doing so, a semantic analysis of a set of sub-technical words will also be carried out in section 4.4 applying Cantos and Sánchez's (2001) lexical constellation model to the analysis of their semantic features and the path followed by this type of words towards specialisation. This method facilitates the better understanding of the nature of the relationship between general and sub-technical words by examining the underlying mechanisms to differentiate these two categories. By delving into the meanings of each of the words present in section 4.4, it can be observed that their semantic features branch out from their nuclear meaning being both related and also

detached from it as regards the acquisition of new senses deriving from the original one, at the core of the constellation.

4.3. A PROPOSAL TO MEASURE THE DEGREE OF SPECIALISATION OF SUB-TECHNICAL VOCABULARY

As often agreed by scholars (Mellinkoff, 1963; Alcaraz, 1994; Tiersma, 1999; Borja, 2000; Orts, 2006), legal English presents serious difficulties not only to the foreign learner but also to non-specialised natives because of its inaccessibility and pomposity. An example of this fact is the *Plain English Campaign*[26] in the UK whose main aim is to simplify legal texts and make them more accessible to the layman. Its convoluted syntactic structures, the use of Latin and Old French phrases, the continuous appearance of archaic terms, or lexical repetition, amongst other features, certainly hinder the understanding of legal texts such as law reports or statutes which the British legal system stems from.

Furthermore, the use of "common words with uncommon meanings", as Mellinkoff (1963: 11) puts it, also contributes to this obscurity. Words like *battery, trial, charge* or *conviction,* which are quite common in general English, acquire new specialised senses in the legal context. They are sub-technical words whose process towards specialisation is hard to quantify.

In spite of that fact, the implementation of Williams' (2001) lexical network model could facilitate considerably the study of shared vocabulary and its context from a quantitative perspective, let us then describe it in greater detail.

---

[26] http://www.plainenglish.co.uk

4.3.1. William's (2001) lexical network model

The lexical network model put forward by Williams (2001) presents a quantitative approach to the study of the context of usage of words by analysing their collocates and co-collocates. The context is extended to word associations beyond the main node (the word under examination) since the networks spread out progressively by also extracting the node's co-collocates and, in turn, the collocates of those co-collocates until the main node is found again to avoid circularity. Williams employs mutual information (Church & Hanks, 1990) as the statistical measure to identify these patterns.

Although the networks must be supervised manually to discard such elements as ungrammatical or unlexical patterns, they provide large amounts of useful and meaningful information which could be applied to the study of sub-technical vocabulary, especially if a comparison is established between a specialised and a general corpus, as will be shown below. In addition, Williams underlines the usefulness of this method in its capability to "reveal patterns that are significant for texts emanating from a discourse community. These patterns may then be used to demonstrate the essential lexis of that community" (Williams, 2001: 5).

One of the problem areas of this approach is the concept of collocation applied to obtain a word's lexical network. His proposal relies on statistical data which, apart from producing large amounts of relevant information, also identifies elements which need to be purged manually. Williams acknowledges the fact that this is a first automatic step which requires the specialist's supervision. Pseudo-collocates which are unlikely to occur together (unlexical collocates), such as *charges review* or *drift sentence* −extracted from *BLaRC* using mutual information−, or which do not abide by the rules of grammar (ungrammatical collocates), for instance, *benefit has* or *regulation see,* must be eliminated by the researcher for a more exhaustive analysis of that context.

Williams reviews the concept of collocation by grouping authors' definitions into four main characteristics which collocations are expected to display. They must be habitual, that is, as Firth remarks, "collocations of a given word are statements of the habitual or customary places of that word in collocational order" (in Williams, 2001: 3). This is why they can be calculated employing statistical measures, following Sinclair's (1991) approach to the concept. However, depending on the tools employed or the measures applied, the number of collocations extracted varies and the label *candidate collocations* must be employed since pseudo-collocates must be manually filtered by the specialist.

Collocations are also described as "lexically transparent", that is, following Cruse "the essential difference between a collocation and an idiom is that in the former each word remains fully transparent whereas in the latter the meaning can no longer be decomposed" (in Williams, 2001:4). However, Williams underlines the fact that such transparency is a matter of degree, owing to the fact that a word could well be part of a collocation and an idiom at the same time.

Another characteristic described by the author is the arbitrariness of collocations across languages. It appears that most collocations cannot be translated literally from one language to another one. However, once more, this is a question of degree, as put forward by Haussman (in Williams, 2001: 4), who allocates collocations along a free-fixed continuum.

Finally, Williams refers to syntactic coherence as a condition for collocational patterns to be considered as such. Kjellmer (in Williams, 2001:4) mentions the relevance of grammaticality as a necessary condition in this respect. Nonetheless, William's perspective on this question is similar to Firth's in that "collocation is

syntagmatic recurrence, which may be described in syntactical terms, but this is not a condition" (Williams, 2001: 4*).*

Williams' concept of collocation could thus be identified with the Birmingham school which relies on statistical measures to extract them. Collocations can thus be calculated applying methods like mutual information (Church & Hanks, 1990) not considering such aspects as grammaticality or lexical transparency in the initial phase to obtain a word's lexical network. The author defines collocation as "the habitual and statistically significant relationship between word forms within a predefined window and for a defined discourse community, expressed through an electronic corpus of texts" (Williams, 2001: 5).

4.3.1.1. Method implementation

To the best of our knowledge, no quantitative method has been designed to date to describe sub-technical words or rank them with respect to their contexts of usage in the specialised or general fields. However, due to the contextual information it is capable of processing, Williams' model might be employed as a useful tool to try and quantify this linguistic phenomenon.

Williams' method takes into consideration not only a word's capacity to generate collocates by itself but also the associations of its collocates and co-collocates, thus increasing the contextual information provided in each case. Therefore, it was employed as the means to obtain the necessary data to calculate a word's sub-technicality level by comparing the networks generated by it both in *BLaRC*, the specialised corpus, and *LACELL*, the general one. The collocate span established was 5 to the left and the right of the node and sub-nodes subsequently. Nevertheless, a >30 collocate frequency threshold was applied to prevent the networks from becoming

unmanageable. Even so, the average number of elements in each network was 2609 for *BLaRC* (294.80 after normalisation) and 596 for *LACELL* (39.83 after normalisation). Additionally, the networks expanded at two levels, that is, they comprised the main node's collocates and the collocates of those collocates (the so-called *co-collocates*) so as to limit their size in a way that the information could be properly handled (according to Williams, they are expected to grow until the main node appears again). Despite that fact, the networks often closed themselves by repeating the main node as co-collocate and therefore not requiring to be re-explored in case further collocational levels had been considered.

Both *BLaRC* and *LACELL* were analysed using *Wordsmith 5.0* (Scott, 2008) employing mutual information in both cases for the results to be more consistent. Function words were filtered out using the function wordlist provided with Heatley and Nation's (1996) software package *Range* (Heatley & Nation, 1996). Following Williams' procedure, ungrammaticality and lexical transparency were not considered for the initial selection of the collocates in each network. On the contrary, once function words had been purged automatically, all the patterns identified by mutual information were included as part of the collocate inventory. Then, the resulting lists of words forming each network were transferred to a spreadsheet and the data obtained was then processed applying step 4 of *Sub-Tech*, the algorithm proposed and explained in section 4.3.2.

4.3.1.2. Sub-technical network sample

In the introduction to this section, sub-technical legal vocabulary was divided into different categories based on both statistical criteria (by comparison between the specialised and general fields) and also semantic ones (depending on the word acquiring

a new meaning in the specialised context and its relation with its original sense). As a result, three different categories were established, namely: 1) Words denoting a legal concept which are frequently used both in the general and specialised fields not changing their meaning in the legal context; 2) words often employed both in the general and specialised fields which change their meaning in the legal context sharing some semantic features with their original meaning; 3) words occurring more frequently in the specialised field than in the general one which change their meaning in the legal environment acquiring a new meaning. Their new meaning is quite distant or completely unrelated to their general sense.

Due to the size of the networks, only one of them will be included in this section to exemplify the method followed to obtain the necessary data to establish a word's sub-technicality level. The word selected is *guilty,* shared both by the specialised and the general fields which does not acquire a new sense in the legal area, thus belonging in category 1). As a matter of fact, due to its widespread use in general English, its sub-technicality coefficient is relatively low in comparison with the rest of words sampled for this study. The value assigned to this word is 2.624 which shows how, in spite of being sub-technical due to its shared character, it is much closer to the general field than to the specialised one standing at the bottom of the sub-technical word list analysed, as will be illustrated below.

This value was obtained by comparing the number and frequency of the collocates in both the specialised and general lexical networks obtained for *guilty.* As illustrated by tables 5 and 6, the specialised network of *guilty* is formed by 589 words while its general network only has 179 elements in it. The differences grow bigger if the figures are normalised since the average number of network elements would be 66.55 in the specialised corpus and 11.96 in the general one. Similarly, the average normalised

frequency of the collocates in both networks is also higher in the specialised corpus reaching 6.87 as opposed to 4.25 in the general one.

Table 5

Specialised lexical network of GUILTY (obtained from BLaRC)

MAIN NODE	MAIN NODE COLLs	CO-COLLOCATES
GUILTY	PLEAD (325)	GUILTY (275), CASE (111), CLAIM (68), APPELLANT (51), FACTS (51), ALREADY (38)
	PLEA (253)	GUILTY (192), BASIS (86), LAW (47), ACCEPTED (39), ENTERED (30)
	OFFENCE (153)	SECTION (303), ORDER (102), MADE (102), ARTICLE (102), RULE (63), PARAGRAPH (46), DECISION (45), ACT (45), AGREEMENT (43), CLAUSE (42), CLAIM (41), COURT (40), REGULATION (38), NOTICE (32), CPR (31), APPEAL (30)
	FIND (78)	TRIBUNAL (388), JUDGE (285), FACTS (197), COURT (162), CASE (142), EVIDENCE (123), CLAIMANT (123), EMPLOYMENT (105), FACT (93), APPELLANT (91), ALSO (88), GUILTY (78), SECTION (72), PARAGRAPH (64), APPEAL (63), LIABLE (62), ARTICLE (60), DECISION (60), LAW (59), DIFFICULT (53), ONLY (50), LORD (48), RELEVANT (47), JUDGMENT (47), MADE (46), FAVOUR (44), RESPONDENT (44), IMMIGRATION (41), CLAIM (39), CIRCUMSTANCES (38), SAID (36), BREACH (36), PART (35), PARA (35), CASES (35), DISMISSAL (34), NECESSARY (34), BASIS (33), FOLLOWING (32), ACT (31), NOW (31), WAY (30), PROVED (30)
	APPELLANT 78)	EVIDENCE (363), CASE (330), APPEAL (297), RESPONDENT (271), MADE (270), BEHALF (259), COUNSEL (252), SAID (244), QC (184), DECISION (158), TRIBUNAL (152), APPEARED (152), INSTRUCTED (142), ALSO (140), GIVEN (128), FACT (113), COURT (112), APPLICATION (110), TIME (109), THEN (103), JUDGE (103), SUBMITTED (102), ENTITLED (96), STATE (94), GAVE (92), FOUND (91), COMPANY (91), COMMISSIONERS (91), TOLD (89), CONVICTED (86), ONLY (86), RESPONDENTS (81), TAKEN (79), PRESENT (78), GUILTY (78), STATED (78), ACCOUNT (78), PERSON (78), TRIAL (78), RIGHT (77), INFORMATION (75), WORK (73), ACCEPTED (72), ORDER (71), SECRETARY (71), CLAIM (70), COSTS (69), SOUGHT (68), POLICE (67), APPEALS (66), FURTHER (66), HEARING (64), HMRC (63), STATEMENT (63), KNEW (63), LEAVE (62), NOW (62), LETTER (61), JUNE (61), GOODS (60), ASKED (60), LORD (59), APPEALED (58), FAILED (58), GROUNDS (58), APRIL (58), REPRESENTATIVE (57), FOLLOWING (57), FIND (57), POSITION (56), VAT (56), POINT (56), DEAL (55), SUBMISSIONS (54), BASIS (54), RECEIVED (54), PART (54), CONSIDERED (54), LIMITED (53), FC (53), ARGUMENT (53), SEEN (52), VERY (52), ISSUE (52), PLEADED (51), PERIOD (51), CLAIMANT (51), CLAIMED (51), SECTION (50), SATISFIED (50), MAKE (50), TRANSACTIONS (49), WROTE (49), QUESTION (49), YEARS (48), RISK (48), JUDGMENT (48), SUBMISSION (48), OCTOBER (47), NOTICE (47), BENEFIT (47), PROCEEDINGS (47), PROVIDED (46), CAUSE (46), VIEW (46), DATE (45), REPRESENTED (45), PAID (44), KNOWN (44),

		SUBMITS (44), USE (44), APPLIED (44), SUPPORT (44), HUSBAND (44), GIVE (44), COUNCIL (44), RELATION (43), PUT (43), ACCEPT (43), PAY (43), HELD (42), DECEMBER (42), ABLE (42), RELIED (42), DIRECTOR (42), APPLY (41), ADMITTED (41), CONSIDER (41), HOUSE (40), CONTENDED (40), TRADING (40), CHARGED (40), SOLD (40), RAISED (40), PROVIDE (40), DENIED (40), SENTENCED (40), NOVEMBER (40), COURSE (40), SHOW (39), REASON (39), WRITTEN (39), WAY (39), EFFECT (39), CLEAR (39), JULY (39), BUSINESS (39), RETURN (39), JANUARY (39), SENT (39), PAYMENT (39), KNOWLEDGE (39), FAMILY (38), SOLICITORS (38), DAY (38), LTD (38), REFERRED (37), CIRCUMSTANCES (37), TAX (37), MARCH (37), OFFICER (37), ARGUED (37), FEBRUARY (37), INFORMED (36), SOLICITOR (36), INDICATED (36), WENT (36), DISPUTE (36), FACTS (36), REQUIRED (36), CONCERNED (36), MET (36), ISSUED (36), DAVID (36), APPEARANCES (35), SUFFERED (35), ENTERED (35), SEPTEMBER (35), GROUND (35), COMMITTED (35), REASONABLY (35), OFFENCE (35), IMMIGRATION (35), INTERVIEW (35), AUGUST (35), RESPECT (35), LLP (34), NAMED (34), USED (34), ADVOCATE (34), PURSUER (34), JURY (34), MOBILE (34), CONCLUSION (34), REASONABLE (34), OWN (34), NUMBER (33), BOUGHT (33), OFFENCES (33), CONDUCT (33), SAME (33), REASONS (33), PREVIOUS (33), CROWN (33), TOOK (33), AGREED (32), CONVICTION (32), SUGGESTED (32), CAME (32), COMMISSIONER (32), DEFENCE (32), SUSPENDED (32), GOOD (31), ARTICLE (31), DATED (31), CHIEF (31), RETURNED (31), PROPERTY (31), INVOLVED (30), ESTABLISHED (30), REQUEST (30), AUTHORITY (30), NEVER (30), CALLED (30), ALLEGED (30), PRODUCED (30), LEFT (30)
	MURDER (65)	GUILTY (65), CONVICTED (54), ATTEMPTED (49)
	DEFENDANT (54)	CASE (170), CLAIMANT (143), MADE (120), BEHALF (115), EVIDENCE (103), COURT (97), ORDER (88), PROCEEDINGS (83), CLAIM (69), PAY (68), SAID (64), PROPERTY (63), TIME (62), ALSO (57), TRIAL (55), APPEARED (54), GUILTY (54), HEARING (54), ONLY (52), CRIMINAL (52), THEN (51), ACTION (50), LIABLE (49), CONVICTED (49), DECISION (48), HELD (47), COSTS (46), DEFENCE (46), PART (45), PARTICULAR (44), JUDGMENT (44), INFORMATION (44), BENEFITED (44), FACT (43), PLAINTIFF (43), OFFENCE (43), ACT (42), APPLICATION (41), BREACH (41), FAILED (41), GIVEN (41), MAKE (40), MB (40), RESPONDENT (40), CLAIMANT'S (39), ISSUE (39), DUTY (38), ENTITLED (38), APPEAL (38), POSSESSION (37), LIABILITY (37), KNEW (37), DRUG (36), CONDUCT (36), SECTION (35), RECEIVED (35), REASONABLE (35), VERY (34), CHARGED (34), OBTAINED (34), FURTHER (33), COMPANY (33), SUBJECT (33), TAKE (33), ALLEGED (32), DECEMBER (32), PRESENT (32), WITNESS (32), QUESTION (32), PARTY (32), DATES (32), TRAFFICKING (31), SOUGHT (31), PERSON (30), GIVE (30), RELEVANT (30)
	COUNT (74)	INDICTMENT (73), GUILTY (42), IMPRISONMENT (32), SENTENCE (31)
	PERSON (34)	ACT (232), DISABLED (202), REASONABLE (197), SECTION (197), TAXABLE (186), RIGHT (156), MADE (155), CONCERNED (147), PROVIDES (138), OFFENCE (130), CASE (123), CONTROLLED (119), ENTITLED (119), RESPECT (111), ONLY (102), QUALIFIED (102), ACTING (99), RELEVANT (98), SUBJECT (95), ORDER (95), RELATION (95), QUESTION (92), PROPERTY (86), APPLIES (85), CONVICTED (85), TIME (84), REQUESTED (84), LIABLE (83),

		PURPOSES (82), COURT (82), ACCUSED (81), WORK (80), ALSO (79), APPELLANT (78), AUTHORITY (78), STATE (78), DISCRIMINATES (78), APPEARED (77), DECISION (75), CLAIMANT (74), SKILLED (74), POSITION (72), MAKE (70), PARTICULAR (69), DETAINED (68), MAKING (67), MEANS (67), GIVEN (65), CIRCUMSTANCES (65), CONTROL (65), SAME (65), PARAGRAPH (62), RIGHTS (61), COMMITTED (60), PART (60), RESPONSIBLE (60), INTEREST (58), CANNOT (58), BEHALF (57), LEGAL (56), SELF-EMPLOYED (55), GROUNDS (55), CONDUCT (55), PERSONS (55), CHILD (54), INFORMATION (54), ARTICLE (54), POSSESSION (54), DESIGNATED (54), APPLICATION (53), EMPLOYED (53), AFFECTED (53), MEMBER (53), APPEAL (52), PROCEEDINGS (52), REASON (52), LAW (52), DISABILITY (52), BENEFIT (51), TREATED (51), YOUNG (50), SEEKING (50), ABROAD (50), ALLEGED (50), HELD (50), APPLY (49), CLAIM (49), EMPLOYER (48), IDENTIFIED (48), GIVE (48), REQUIRED (48), EFFECT (47), EVIDENCE (47), FAMILY (47), CHARGED (46), SAID (45), TAKE (44), BODY (44), LIBERTY (44), OFFENCES (44), PROVIDE (43), PURPOSE (43), HEARING (43), INVOLVED (43), GRANTED (43), SUBSECTION (42), LIKELY (42), CRIMINAL (42), PARTY (41), REGULATION (41), COMMITS (41), LIFE (41), PROVIDED (41), DUTY (40), REASONABLY (40), ASYLUM (40), REGISTERED (39), CONSIDER (39), PAY (39), DIRECTOR (38), OPINION (38), COURSE (38), EMPLOYMENT (38), SITUATION (37), HOMELESS (37), TAKEN (37), PRESENT (37), COMMITTING (37), INCLUDE (36), KNOWLEDGE (36), REGULATIONS (36), DIRECTLY (36), UNITED (36), APPOINTED (36), IDENTIFY (36), NAME (36), NOTICE (36), CARE (35), RISK (35), ACTS (35), OTHERWISE (35), FACT (35), USE (35), IMMIGRATION (35), SOUGHT (34), FOLLOWS (34), SUPPLY (34), FAVOURABLY(34), GUILTY (34), RESPONDENT (34), NEED (34), NATURAL (33), WORKER (33), DETENTION (33), TREATS (33), VAT (33), ACTUALLY (32), THEN (32), ACTION (32), ENGAGED (31), CASES (31), APPLICANT (31), DECEASED (31), COUNTRY (31), FIT (31), EXTRADITION (31), DEFENDANT (30), PUBLIC (30), COMPANY (30), KNOWN (30), TREATMENT (30), SECURITY (30), CONNECTED (30)
	CHARGE (34)	CRIMINAL (60), SECTION (45), CASE (45), EQUITABLE (44), SERVICE (42), CREDIT (40), FIXED (39), TAX (36), LEGAL (35), GUILTY (34), RESPECT (33), SUBJECT (31)
	ENTER (32)	AGREEMENT (138), CONTRACT (63), TRANSACTIONS (56), PARTIES (43), TIME (42), KINGDOM (37), UNITED (37), APPELLANT (35), JUDGMENT (34), GUILTY (32), PLEA (30)
	MASLAUGHTER (31)	GUILTY (31)

Table 6

General lexical network of guilty (obtained from LACELL)

MAIN NODE	MAIN NODE COLLs	CO-COLLOCATES
GUILTY	FEEL(77)	MAKE (429), PEOPLE (182), VERY (165), BETTER (124), JUST (123), NEED (109), KNOW (105), REALLY (103), NOW (100), GOOD (96), WAY (86), WELL (86), COMFORTABLE (79), GUILTY (77), QUITE

		(64),THINK (64), CONFIDENT (62), SORRY (60), THEN (59), TOO (59), TIME (59), ONLY (58), WOMEN (58), SAY (57), GOING (57), SAID (56), RIGHT (55), BEGAN (54), BIT (54), ABLE (49), FREE (49), THINGS (48), WANT (48), LOOK (48), SEE (44), GO (44),OWN (42), ALSO (41), OFTEN (41), IMPORTANT (40), MEAN (39), CHILDREN (39), SURE (38), NEVER (38), AGAIN (37), PERSON (37), ALWAYS (37), TIRED (37), SAME (36), LITTLE (35), STRONGLY (35), MAKING (35), SAFE (35), MAN (34), HELP (33), SICK (32), OLD (31), BAD (31), COME (31), SOMETIMES (31), PERHAPS (31), GREAT (30), DIFFERENT (30), PARENTS (30), UNCOMFORTABLE (30), LOT (30)
	FIND(69)	DIFFICULT (276), WAY (246), PEOPLE (229), VERY (175), ONLY (147), HARD (144), THEN (135), ER (127), TIME (118), TRY (118), WORK (109), NEW (108), THINK (105), HELP (102), ERM (102), WELL (94), WANT (91), ALSO (90), NEED (89), JUST (88), TRYING (85), NOW (84), RIGHT (83), GO (80), KNOW (75), WAYS (75), OFTEN (73), ABLE (72), EASY (71), QUITE (69), PLACE (68), ALWAYS (67), GOING (67), NUMBER (64), GOOD (64), THINGS (63), AGAIN (63), HERE (61), WOMEN (61), TOO (60), USEFUL (59), USE (58), SAID (57), SEE (53), FORMULA (52), NEVER (51), ACTUALLY (51), BEST (51), EASIER (50), OWN (50), EXPECT (50), LITTLE (49), ELSE (48), WORDS (48), BOOK (47), REALLY (46), HOPE (46), INTERESTING (46), LOCAL (45), BIT (45), HELPFUL (44), CANNOT (43), POSSIBLE (42), BETTER (42), SOLUTION (41), FOLLOWING (41), PROBABLY (41), MAN (41), AREA (40), JOB (40), COME (40), INFORMATION (40), WORLD (39), SAY (39), LONG (39), MAKE (39), CHILDREN (39), DAY (39), STUDENTS (38), LIFE (38), UNABLE (38), MEN (38), LOOK (37), SAME (37), PARTICULAR (36), OLD (36), YOUNG (36), SURPRISED (36), KIND (35), ANSWER (35), FACT (35), LIKELY (35), LOT (34), TAKE (34), LATER (34), EXAMPLE (34), SOMETIMES (33), FOOD (32), GIVE (32), MONEY (32), PROBLEMS (31), SORT (31), IMPOSSIBLE (31), USED (31), MEAN (30), DIFFERENT (30), PERHAPS (30)
	PLEAD(43)	GUILTY (43)

4.3.2. *Sub-Tech*: the algorithm

After calculating the lexical networks of several technical, sub-technical and general words, a direct observation of the information provided was carried out leading us to the formulation of a method to attempt to objectively determine a word's sub-technicality level based on a comparison between the number and frequency of the collocates and co-collocates found in those networks both in the specialised corpus, *BLaRC*, and the general one, *LACELL.*

An algorithm could therefore be proposed including all the steps followed towards the ranking of this type of vocabulary. The algorithm, owing to its major objective, was called *Sub-Tech*. It goes as follows:

Step 1: Identification and extraction of the specialised single and multi-word terms in *BLaRC* applying both Drouin's (2003) *TermoStat* and Nazar and Cabre's (2012) *Terminus 2.0.*

Step 2: Manual extraction of those words whose level of specialisation (according to the methods selected) was not excessively high being shared both by the general and specialised fields.

Step 3: Application of Williams' (2001) lexical network model to the list of words selected both in the specialised and general corpora with the aim of comparing results.

Step 4: Implementation of the formula presented below in order to rank sub-technical terms along a continuum of specialisation.

4.3.2.1. Ranking method: development and justification

A lexical network R_i illustrates the relationship between a given word w_i and its collocates and co-collocates $\{w_1, w_2, \ldots, w_N\}$ in a given context. These relationships are determined by the number of times $\{\mu_1, \mu_2, \ldots, \mu_N\}$ such elements co-occur in that context. Depending on the corpus the network was obtained from, a distinction will be made between the specialised one based on the legal corpus R_i^T and the one based on the general corpus R_i^G.

Having empirically examined the data available, it was observed that, in general, when employing the legal English corpus as the source to obtain the lexical networks from, specialised terms as well as sub-technical ones tended to attract a greater number

of collocates and co-collocates also displaying higher frequency counts[27]. On the other hand, words which belong to general usage showed a smaller number of collocates and co-collocates in the same context having considerably lower frequency counts.

On the contrary, as regards the networks based on the general corpus (*LACELL*), data behaved conversely, that is to say, specialised terms were associated with a smaller amount of elements and displayed lower frequency counts, whereas general usage words tended to co-occur with a greater number of words much more frequently. Table 7 illustrates this tendency. The figures in table 7 have been normalised for the values to be comparable because of the different size of both corpora.

Table 7

Number and frequency of the collocates and co-collocates in the lexical networks analysed

Word	Specialised collocates/ co-collocates (normalised)	General usage collocates/co-collocates (normalised)	Normalised Frequency *BLaRC*	Normalised Frequency *LACELL*
PURSUANT	404.40	0	10.34	0
ESTOPPEL	114.57	0	8.65	0
LIABILITY	421.69	0	8.20	0
BATTERY	27.57	0.73	7.89	2.27
CONVICTION	281.35	1.33	10.41	3.23
SENTENCE	491.25	1.53	9.50	2.98
DISMISS	338.64	3.20	10.06	3.81
SOLICITOR	159.77	0.33	8.23	2.39
RELIEF	184.18	6.08	9.88	4.45
TRIAL	666.66	2.33	9.22	3.84
LEGISLATION	246.44	39.7	9.23	4.2
WARRANT	30.39	1.60	7.91	3.01
PARTY	708.36	274.13	9.22	4.73
CHARGE	167.68	64.77	9.08	4.89
COMPLAINT	180.22	18.18	8.79	4.70
OFFENCE	522.93	28	8.91	5.03
GUILTY	66.55	11.96	6.87	4.25
EAT	0	2.20	0	3.27
BLUE	0	13.43	0	3.52
MORNING	0	268.36	0	4.94

---

[27] Frequency counts refer to the multi-word units formed by the node and its collocates and the sub-nodes and their collocates (the co-collocates). They indicate the number of co- occurrences of these elements in each corpus.

As it can be observed from table 7, the words *pursuant, estoppel* and *liability* are clear examples of highly technical terms which do not generate any collocates in the general corpus applying the >30 frequency threshold established in this study. Conversely, *eat, blue* and *morning,* general words employed in everyday language, have no collocates in the specialised corpus under the same circumstances. The rest of them are sub-technical, either because they are shared by the general and specialised areas of language without changing their meaning (*guilty; warrant; legislation; solicitor*) or because they acquire a technical meaning in the legal context (*trial; relief; sentence; conviction*). All of them have a greater number of collocates and co-collocates in the legal corpus. On average, the sub-technical words analysed in this section generate 258.29 more collocates and co-collocates in *BLaRC* than in *LACELL.* In addition, the average frequency of these collocations is higher for all of them in the specialised corpus (5.1 points higher in *BLaRC* on average) except for *eat, blue* and *morning,* which behave conversely due to their highly general character.

The formula below was thus designed to try and measure this phenomenon:

$$\bar{\mu}_\iota = E(\mu_j) = \frac{\sum_{j=1}^{N} \mu_j}{N}$$

Where $E(.)$ represents the expected value (average). This method allows us to establish the degree of concentration of the elements included in a lexical network rather straightforwardly.

The coefficient of sub-technicality of a given word $ST(w_i)$ will thus be calculated by subtracting the average frequency of the collocates and co-collocates in the general usage network $\overline{\mu_\iota^G}$ from the same parameter in the specialised one $\overline{\mu_\iota^T}$. Both values must be normalised by dividing them by the total number of elements in each corpus, that is:

$$ST(w_i) = \frac{\overline{\mu_i^T}}{|C^T|} - \frac{\overline{\mu_i^G}}{|C^G|}$$

Where $|C^T|$ and $|C^G|$ represent the number of elements in the specialised and general corpora respectively. This normalisation is necessary so as to obtain a coherent value due to the different size of both corpora. The average frequency values in each corpus can only be compared if they are normalised since *LACELL* (14.6m words) is almost twice as big as *BLaRC* (8.85m). In order for the figures obtained to be manageable, $|C^T|$ and $|C^G|$ were expressed in millions of words.

The coefficient proposed above was not delimited either superior or inferiorly, therefore, the values obtained cannot be studied in isolation but rather as part of a continuum of technicality where words will tend towards one or the other end. Those words displaying a higher value will be considered more technical than those showing a lower one. Figure 16 exemplifies this continuum.
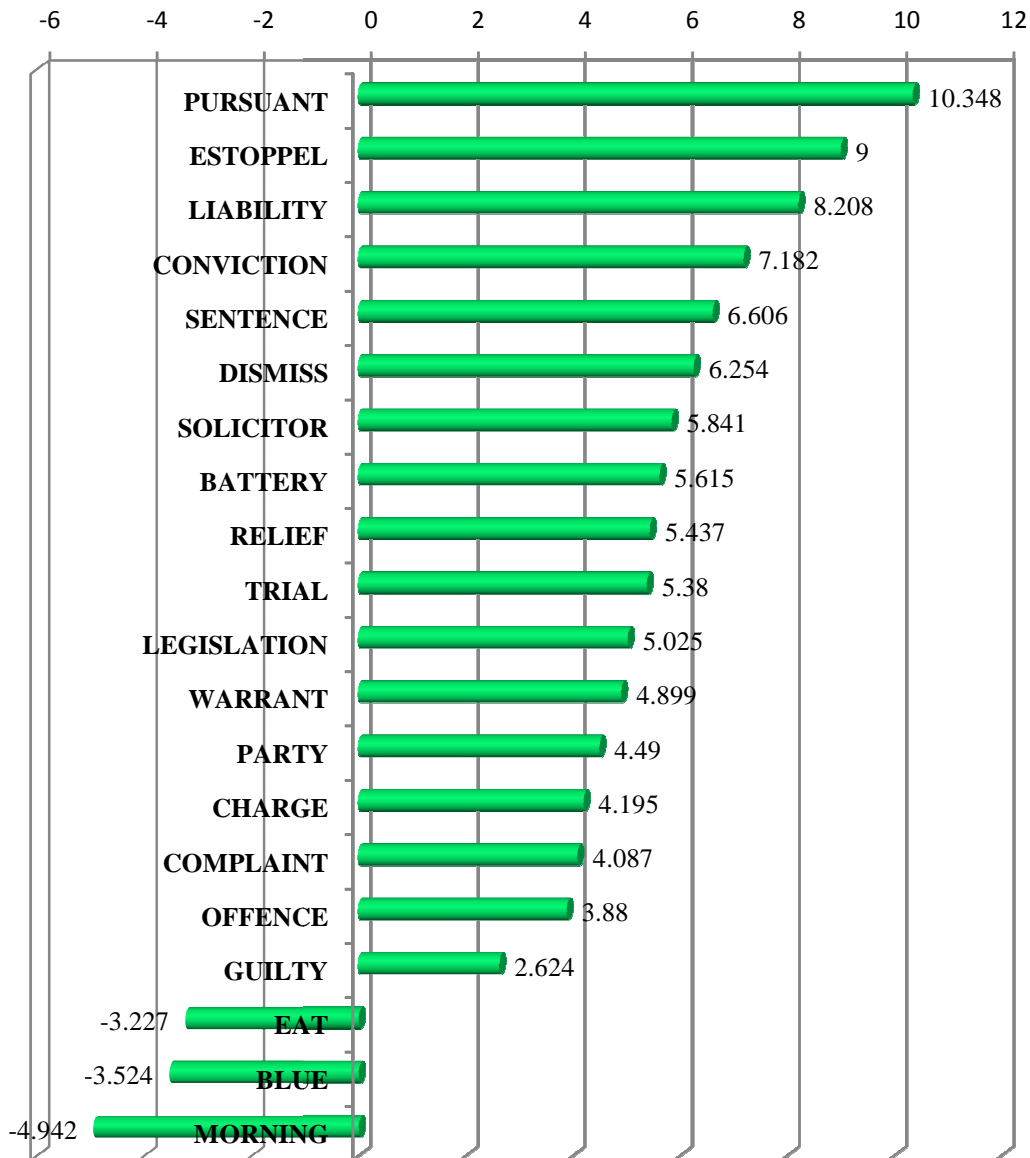
Figure 16. *Sub-technicality coefficient*

As illustrated in figure 16, *pursuant*, *estoppel* and *liability,* the most specialised

words in the group which do not generate any collocates in the general corpus, display

the highest coefficients[28] (10.348, 9, and 8.208 respectively) standing at the higher end

of the continuum of specialisation, whereas *eat, blue* and *morning* (whose technicality

---

[28] The coefficient is labelled as sub-technical in figure 10 due to its main aim, that of measuring a word's sub-technicality level, however, we cannot refer to the concept *sub-technicality* when testing the method on highly specialised terms or general words.

coefficients are -3.227 -3.524 and -4.942 respectively), appear at the opposite end owing to their highly general character. They were employed to test the method showing that, while the rest of words are distributed between these two extremes, they would gather at opposite ends of the continuum as a result of their lack of collocates in either the specialised or the general networks once the algorithm *Sub-Tech* was fully implemented.

Nevertheless, for the method to work properly, the statistical data of the words selected must be significant. If it was not, in spite of its robustness, mutual information, as well as other statistical measures, would produce misleading results. Therefore, if step 4 of the algorithm was applied to a word like *lessor,* which occurs only 50 times in *BLaRC* and none in *LACELL* not generating any significant collocates above the >30 frequency threshold established, it would not be located within the highly technical term range (where it belongs) but rather within the general one. Even so, this method has not been designed to measure that type of words but rather those which tend to one end of the continuum or the opposite one due to the number and frequency of the constituents of its lexical network in both corpora.

As regards the rest of the words, the sub-technical set, they are distributed along the continuum depending on the comparison established between the number of elements in their specialised and general networks. There seems to be no correlation between the semantic classification of these words offered in section 4.2. and their position along this continuum. As a matter of fact, those words belonging to semantic category 1, which do not change their meaning either in the specialised or general fields, display completely different coefficients ranging from 5.841 for *solicitor* (considerably close to the most specialised terms) to 2.624 for *guilty* (the lowest ranking sub-technical word from the set).

On the other hand, those words which acquire a specialised sense in the legal context related to its general meaning (semantic category 2) do not distribute evenly along the continuum since *offence, complaint,* and *charge* differ 4/5 points from the most technical word group and about 7 from *eat,* the highest ranking general word. However, they stand at the very bottom of the list together with *guilty,* from semantic group 1. On the contrary, *trial, relief,* and *battery,* other members of this group, appear in a middle position (their coefficients are 5.38, 5.437, and 5.615 respectively), while *dismiss* or *sentence* are much closer to the technical set at less than 1 point of distance from *liability* (8.208), a technical term.

Likewise, those words belonging to semantic category 3, whose specialised sense is far from its general one, are also distributed along the continuum irregularly although they appear to be more distant from the general set than group 2. Actually, *party* and *warrant* (with 4.49 and 4.899 coefficient respectively) stand 7 points above *eat,* that is, 2.3 points more than the lowest ranking words in group 2. As for *conviction* (7.182), it remains much closer to the highly technical set being the most "technical" of the sub-technical word sample examined in this section.

4.4. CANTOS AND SÁNCHEZ'S (2001) LEXICAL CONSTELLATION MODEL: A SEMANTIC ANALYSIS OF SUB-TECHNICAL WORDS

In spite of having proposed a method to attempt to measure the degree of specialisation of sub-technical vocabulary, much still remains to be said about the manner in which this type of words become specialised terms. As Rea and Sánchez (2010) assert, technical language requires the creation of new words to name new concepts related to a given subject field. The mechanisms to create new words are varied: word coinage, derivation or borrowing. However, polysemy is an economic way to solve the problem

by assigning to an already existing form of language or *denotandum* a new sense or *denotatum* thus requiring a smaller effort on the part of the speaker. This is the means by which sub-technical words acquire new meanings in the sense that they activate them when inserted in a technical context.

Cantos and Sánchez (2001) present a novel approach to lexical analysis which could be applied to the understanding of the path followed by semi-technical words towards specialisation: the *Lexical Constellation* (LC) model. It consists in studying "the way words socialise with other words, forming complex network-like structures or units" (Cantos & Sánchez, 2001: 200) which are hierarchically organised displaying semantic dependencies according to their rank within the network. Actually, these constellations work similarly to a star system where planets orbit around a central star, the node, which attracts them, being connected, in turn, to other star systems forming constellations, hence their name.

The LC model manages to overcome such limitations as the establishment of the optimum span to filter the number of collocates to be considered for analysis by setting the sentence as the limit for such span. Not only does this model provide information about the most significant collocates of a given word but also about the hierarchical relationship between a word and the constituents of its constellation. Furthermore, it manages to represent those relationships in a visual and multi-dimensional way facilitating to a great extent the understanding of the dependencies existing amongst the constituents of each LC.

As commonly agreed by researchers, the different senses of a word are neither fully transparent nor clear-cut (Cruse, 2000; Almela, 2006, Kilgariff, 2006; Rea and Sánchez, 2010) therefore, they could be interconnected forming a network where the nucleus is added new features which stem from it leading to the acquisition of new

meanings. Consequently, the purpose of this section is to try and demonstrate how the general and specialised fields interact and how specifity is generated through such interaction. The LCs can explain sub-technical terms through the analysis of the semantic hierarchies existing amongst the general and specialised semantic features of these terms and their dependencies in a very clear and visual manner, showing that "semantic bonds and the configuration of conceptual associations ... [are] multi-dimensional" (Sánchez et al., 2010: 142).

In an attempt to suggest the LC model as one which can provide reliable contextual information for automatic word sense disambiguation (WSD), Sánchez et al. (2010) also suggest a parallelism between the way LCs work and the inter-connections of neurons in the human brain. Synapses allow neurons to send electrical or chemical signals to each other which could be translated into information storage, amongst many other functions. In doing so, they stand at the centre of a network communicating with other neurons in other networks, and so forth. Lexical constellations behave similarly in the way that the main node relates to its collocates and these collocates to other collocates, inter-connecting its constituents and being organised in a hierarchy of semantic dependencies. In addition, WSD processes are conditioned by the information provided by the context in a way similar to the description of the concept of *activation* in connectionist theory:

> "The identification of a specific disambiguation path depends on the input received from contextual elements, very much in the way connectionist theory refers to 'activation'. The real biological model is based on a network in which all units (neurons and synapses) are candidates for activation at any time. Activations have the potential to trigger an action, or to initiate a process in which many other neurons and modules may be implicated and participating (spreading activation).

This may be an inspiring model for the disambiguation process of senses through contextual semantic features" (Sánchez et al, 2010: 144).

Rea and Sánchez (2010) illustrate these semantic dependencies with the LC of the word *heart,* as shown in figure 17[29]. In it, shared and new semantic features interrelate forming a network around the word's base meaning: "physical organ in persons/ animals". This original meaning, as attested by Sánchez et al. (2010), expands into three main directions acquiring three new semantic senses related to the original one, namely, "being the central part of something"; "having the shape of a heart"; and "storing thoughts, emotions or feelings". The authors consider eight different senses within the LC of *heart* which derive either directly or indirectly from the basic one, which go as follows:

"1. The hollow, muscular organ in a vertebrate animal that receives blood from the veins and pumps it through the arteries by alternate dilation and contraction.

2. The human heart considered as the central part or core of something.

3. The human heart considered as the essence of something.

4. The human heart considered as the centre or source of emotions, personality attributes, etc.

5. Any of the various humane feelings like compassion, love, devotion, enthusiasm, etc.

6. Courage or bravery as rooted in the heart.

7. Something like a heart in a shape; conventionalised design or representation of a heart shaped like this.

8. A red figure of a heart stamped on a playing card, or a playing card marked with a conventionalised figure of a heart" (Rea and Sánchez, 2010: 113).

---

[29] Taken from Rea and Sánchez (2010: 112).
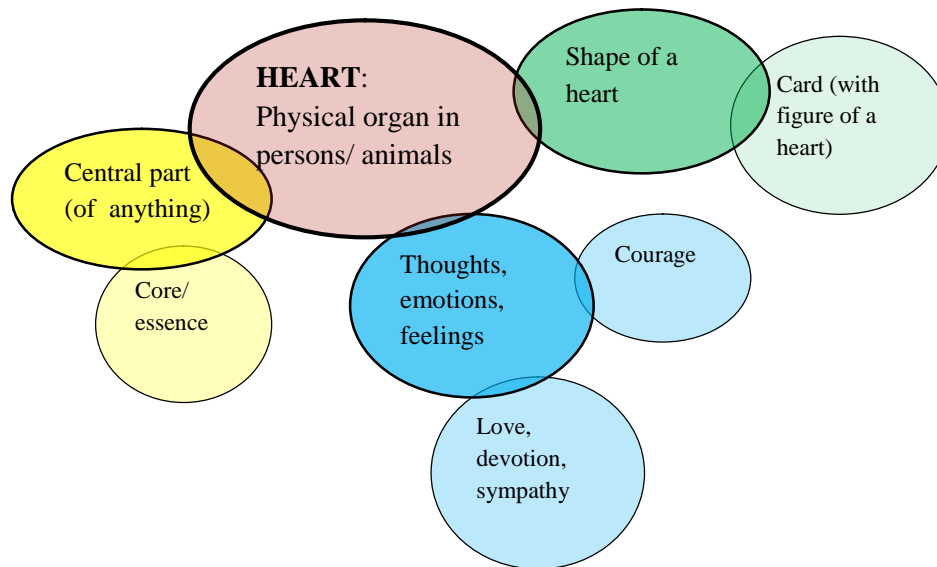
Figure 17. *LC of HEART*

Figure 17 reveals how the three main senses of the word *heart* stem directly from the original meaning, "physical organ in persons or animals" (meaning 1 in the list), having a lower rank than the original one because of their dependency on it. Likewise, the rest of its semantic features derive from these three therefore standing at a lower level than the ones they originate from. These hierarchical dependencies would continue developing if further senses of the word could be associated to them. As it can be observed from the LC of *heart,* the lower the rank of the semantic senses, the more distant they are amongst themselves, as pointed out by Rea and Sánchez (2010). As a matter of fact,

> "if the process of addition of meanings went on, a stage could be reached in which the end-meanings would be so distant from each other that their common source could be hardly visible or recognisable" (Rea & Sánchez, 2010: 113).

This section will therefore be devoted to the application of the LC model to the analysis of the semantic features of a sample of sub-technical nouns in the legal field, namely, *trial*, *charge* and *battery*, following the procedure presented in Rea and Sánchez's (2010) study of the words *chip, bus* and *hub* in telecommunications English.

4.4.1. *Trial*

As illustrated in the *Oxford English Dictionary* (2002)*,* the word *trial* was first recorded in the mid 15th century originally meaning "the act or process of testing". Its legal sense is documented for the first time in 1570, one century later. *Trial* is a noticeably common word in English being present within the most frequent 2,000 word families of *BNC* and the first 1,000 of West's (1953) *GSL* and Coxhead's (2000) *AWL.* Therefore, its usage is widespread in the general field. However, if a comparison is established between its normalised frequency value in *LACELL*, 54.31, and the same parameter in the specialised corpus (*BLaRC*), 489.03, it seems quite clear that, in spite of its extensive use in general English, its presence in the legal context is much more significant.

Nevertheless, after observing the collocates it generates in *LACELL* applying MI as the measure to obtain them, words like *court*, *judge* or *murder* appear amongst its most frequent collocates, attesting that its legal meaning is also considerably frequent in general English. The following concordances obtained from *LACELL* illustrate this fact:

> ... The jury in the James <u>murder trial</u> at Preston Crown Court has been read a statement by a train driver who described how two days after James's disappearance, he saw human remains on the railway line at Walton ...
>
> ... The federal government contends that the <u>trial judge</u> erred on 38 points of his ruling ...

Using the taxonomy provided in section 4.2, *trial* could be identified with category 2), that is, "words often employed both in the general and specialised fields which change their meaning in the legal context sharing some semantic features with their original meaning", as it will be shown below, its legal meaning stems directly from its original one denoting a type of proof or test before a judge.

The value assigned to *trial* after implementing the ranking method presented in section 4.3.2. is 5.38, only 0.72 points below the average value for the set of words studied in that section, which corroborates the fact that *trial* stands at an intermediate position between highly technical terms like *pursuant* or *estoppel* and general words such as *morning* or *blue* owing to the number and frequency of the collocates it generates in the general and specialised corpora.

The *OED* offers the specialised meaning of *trial* as its first sense (the *Merriam Webster Dictionary* offers it as the second option), let us then examine its main definitions to try and obtain the LC of the sub-technical term *trial*:

1) Examination of evidence and applicable law by a competent tribunal to determine the issue of specified charges or claims.

2) The act or process of testing, trying, or putting to the proof: *a trial of one's faith*.

3) An instance of such testing, especially as part of a series of tests or experiments: *a clinical trial of a drug*.

3) An effort or attempt: *succeeded on the third trial*.

4) A state of pain or anguish that tests patience, endurance, or belief: "the fiery trial through which we pass" (Abraham Lincoln).

5) A trying, troublesome, or annoying person or thing: *The child was a trial to his parents.*

6) A preliminary competition or test to determine qualifications, as in a sport.

Figure 18 illustrates the lexical constellation of *trial:*
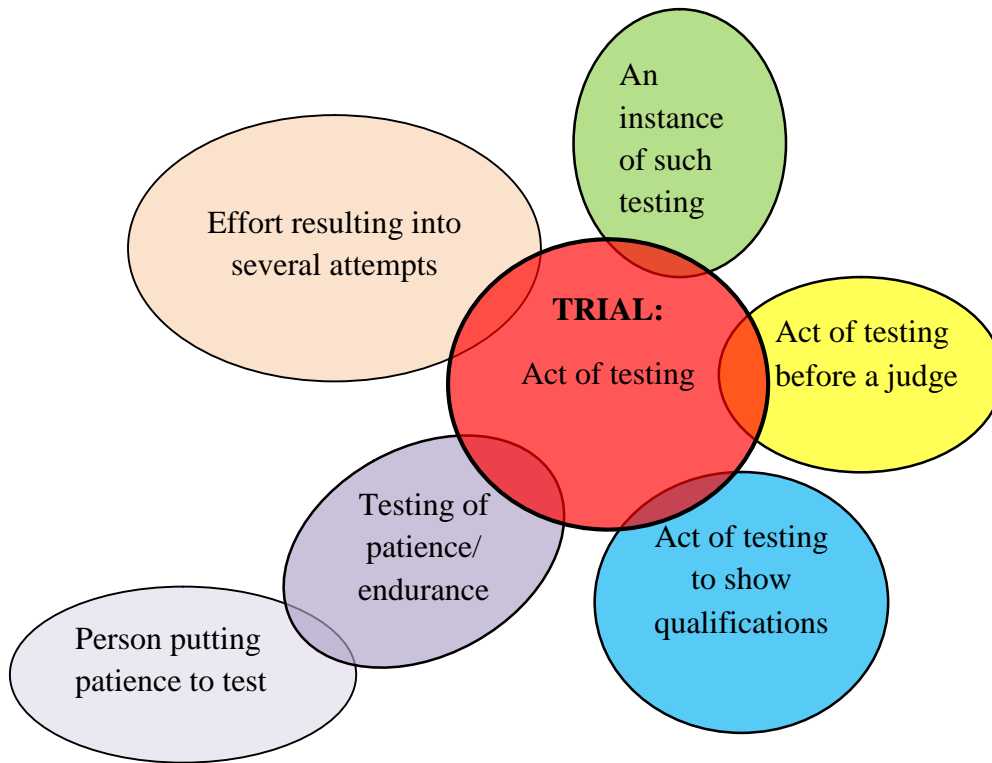
Figure 18. *LC of TRIAL*

The general or basic definition of *trial* could be considered the original one offered by the *OED* (2002), namely, "the act or process of testing, trying, or putting something to the proof", although it is not the first meaning provided. Its greater frequency as a legal term both in the general and specialised fields might account for that fact. As exemplified by figure 18, the legal sense of this word is added to the general idea of testing or proving something since, broadly speaking, a *trial* is a process whereby the validity of the evidence given at court must be proven and, using the legal background applicable, a decision must be made by the tribunal in charge of hearing the case. The concordance below demonstrates this usage:

The court has active case management duties. This includes "deciding promptly which issues need full investigation and <u>trial</u> and accordingly, disposing summarily of others" (*BLaRC*).

Apart from its legal meaning, other four different semantic senses stem from this basic sense of *trial,* as illustrated by its LC in figure 18. A *trial* can be regarded as a "preliminary test to determine the qualifications" of a group of candidates, that is, they are tried to evaluate their potential to suit the purpose of the task they have been assigned or the post they may have applied for appropriately. Therefore, this sense may be regarded as a type of testing and could thus be closer to the legal meaning of *trial*. Instead of testing or proving evidence, the candidates qualifying for a competition, a contest, etc. have to prove their worth in front of a jury or a judge (figuratively speaking) who will make a decision as regards their suitability, for instance, for a given job.

Also within the general field, a *trial* is defined as "a part of a series of experiments or scientific tests", keeping the original sense of the word which highlights the very process of testing or trying something. As a consequence, the word *trial* could be interpreted in this case as "an instance of the act of testing". This semantic sense is often used within the area of clinical testing. When a new drug or substance is discovered, it is tried on a group of patients or volunteers in order to observe its benefits, effectiveness or side effects.

The concordances below found in *LACELL,* the general corpus, illustrate these two different semantic senses:

171

- All morning the world's top cyclists will have been reconnoitring the course, selecting their gear ratios and the most suitable of their many bikes, looking at the route for the best way through corners, testing the strength and direction of the wind - looking for anything that will give them that little extra in this first <u>trial of strength</u> ...

- Whether preventing the formation of uric acid would be beneficial, or whether the accumulation of unconverted xanthine would do more harm than good was open to question, and only to be settled by cautious <u>clinical trial</u>.

The third semantic feature connected with the base meaning of *trial* is associated with the concept of *effort,* adding a rather figurative nuance to the idea of trying or testing. In this case, a *trial* is understood as a succession or a series of attempts to perform a task that poses certain difficulties for its accomplishment, for example: "she succeeded after the third trial", as illustrated by the *OED* (2002). Being less literal than the previous senses of this word, this idea of *trial* could somehow be interpreted as the testing of someone's perseverance to achieve a goal and could thus be related to its fourth meaning, as shown below.

As far as its four definition is concerned, the original idea of testing something for proof can also be associated with "a state of pain or anguish that tests patience ... or belief", as shown in the quotation by Abraham Lincoln provided by the *OED:* "the fiery trial trough which we pass". This kind of test is similar to the previous one as both could be understood as a test of someone's perseverance o endurance. In this case, this test will demonstrate somebody's willpower to overcome difficulties to prove, for instance, their religious faith, whereas in the previous case, *trial* rather denotes a succession of attempts to accomplish a given objective showing someone's insistence on succeeding.

172

Connected to this last interpretation of the word *trial,* though not deriving directly from its base meaning and thus displaying a lower rank within the LC structure, there is the idea of a person or thing which can put somebody's endurance to test being capable of making them lose their temper as in the example "the child was a trial to its parents".

Summing up, as demonstrated by the LC in figure 18, all the semantic senses of the word *trial,* except for one, derive directly from its basic meaning, that is, they stand at a higher level within the LC semantic network, showing a semantic proximity amongst them, as already attested. However, its fifth sense, "a trying, troublesome, or annoying person or thing", displays a lower rank in the LC structure due to its more indirect dependency on the base meaning. Owing to that fact, its connection with the other different meanings of the word is harder to establish showing almost no resemblance, for instance, with its first and third definitions respectively: the "act of testing before a judge" or an "effort resulting into several attempts", which originate directly from the base meaning of this sub-technical term. This is so due to its position within the semantic hierarchy depicted by the constellation of the term which reveals the connection of this feature with the nucleus through an intermediate sense standing at a second level, that of "testing of patience or endurance".

4.4.2. *Charge*

The origins of *charge* date back to the 13th century deriving from the Old French word *charger,* "a load, a weight". According to the *OED,* there is no documented use of this word with a legal sense until the middle of the 15th century when it was also employed to refer to a formal accusation for having committed a crime. It will not be until the beginning of the 16th century that *charge* starts being employed to refer to a "pecuniary

burden" or "cost". Later in the 18th century, with the discovery of electricity, it will be added a new sense to denote the electrical load necessary for a device to work.

Concerning its usage in the general field, it can be found amongst the most frequent 1,000 word families of *BNC,* West's (1953) *GSL* and Coxhead's (2000) *AWL.* It is more frequent than *trial* in *LACELL,* the general corpus, displaying a normalised frequency value of 137.19 as opposed to 164.40 in *BLaRC*. Judging by the figures, it seems that the use of *charge* in the general corpus is much more extended than *trial* (with a general sense) while it occurs three times less than the latter in the legal corpus.

Nevertheless, frequency is not indicative of the type of meaning sub-technical words may acquire depending on their context, since the specialised sense of a word like *trial* is more frequent than other general senses of this word not only in the legal corpus but also in the general one, hence the importance of examining the collocates of the words under examination. Whereas 58.33% of the main collocates of *charge* found in the legal corpus are either specialised terms or sub-technical words (*criminal, case, equitable, legal)*, all the collocates generated by it in *LACELL* belong to the general field. This information might be a clear hint at its specialisation occurring mostly in the legal context.

The word *charge* also belongs to sub-technical word category 2) owing to the fact that its legal meaning can be related to its basic one relatively easily. Its sub-technicality coefficient is slightly lower than *trial,* 4.195. Therefore, it can also be allocated in a middle position along the sub-technicality continuum described in section 4.3. It only differs 0.5 points from the average value for the whole list of words examined in this section. In spite of the considerable difference between *trial* and *charge* as regards their immediate collocates, expanding the context of study by

implementing William's (2001) lexical network model shows that, from a quantitative perspective, *trial* and *charge* are quite similar.

Charge is a very rich word from a semantic point of view. The fact that it can be used as a verb and a noun increases the amount of semantic features attributed to it. This is why, in order not to make the analysis too complex, it will only be considered as a noun herein. Its main semantic features can be grouped into eight main senses considering "a load, a weight" as the original one or base meaning. Only five of these senses can be directly connected with the base meaning of the word, while the other three originate at a lower semantic level in the LC structure, as shown in figure 19. The different semantic senses of *charge* found in the *OED* can be grouped as follows:

1) A weight; a burden; a load: *a freighter relieved from its charge of cargo.*

2) Expense; cost; price asked for something.

3) A quantity of explosive/ gunpowder to be set off at one time.

4) An assigned duty or task; a responsibility: *The commission's charge was to determine the facts.*

5) A claim of wrongdoing; an accusation: *a charge of murder.*

6) A debt or an entry in an account recording a debt: *Are you paying in cash or is this a charge?*

7) The amount of electric energy loaded in a battery: *the electricity charge didn't suffice for the phone to work properly.*

8) A rushing forceful attack: *the charge of a herd of elephants.*
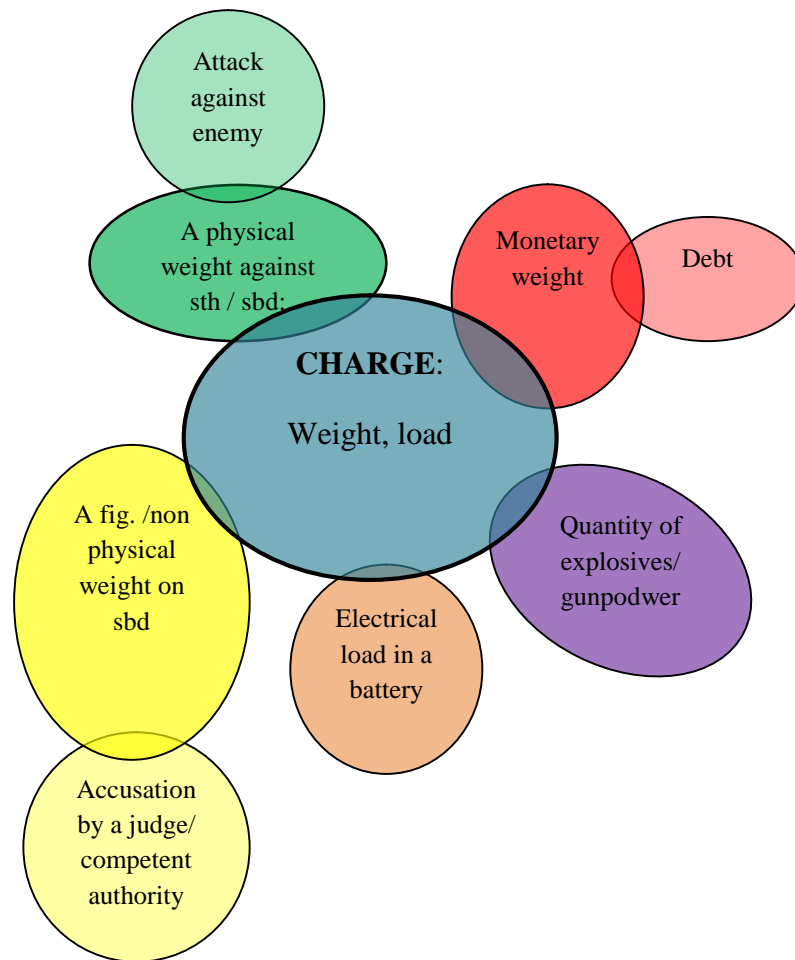
Figure 19. *LC of CHARGE*

As illustrated by figure 19, the basic meaning of *charge* expands into five main directions which result into a higher semantic level within the LC network. The original "weight" or "load" denoted by it can be interpreted as a "monetary weight"; "an electrical load in a battery"; "a quantity of explosives"; "a physical weight against someone" and "a non-physical weight on someone". Except for the first and last senses of *charge* shown above, the other three could be understood as rather literal interpretations of its base meaning owing to the fact that they can somehow be physically felt or measured. In fact, an amount of explosives ready to detonate is

something that can be touched or weighed, if necessary, as well as the electrical load in a battery which generates energy that turns into light, motion, etc.

Regarding the fourth meaning in the list above, the physical charge entailed by this word in an example like "the troops completed the charge against the enemy lines" could be read as one which is physically launched or activated against others, that is, an "attack against the enemy", which is the semantic feature that stems from that "physical weight against somebody". This "attack" would therefore stand at a secondary level in the LC of *charge*.

Nevertheless, a "monetary weight" and a "non-physical weight on somebody" cannot be understood *verbatim* but rather as figurative weights that someone has to bear. The cost assigned to goods might be interpreted as a "load" attached to those goods which the buyer must "carry" to purchase them. Directly deriving from this semantic sense, there is the concept of "debt" whose level of abstraction is higher due to its indirect relationship with the original meaning of the word *charge,* thus holding a lower semantic rank in the LC, as shown in figure 19. A debt might be regarded as a weight on someone's conscience which must be born until it is settled.

On the other hand, the legal sense of *charge* is closely linked to the "non-physical weight on somebody". A legal *charge* could be interpreted as the burden which is placed on somebody's shoulders when he/she is formally accused of a crime by a legal authority. As well as the "attack against the enemy", this meaning of *charge* has a lower semantic rank than the one it originates from implying a further level of abstraction in the LC semantic network and an indirect dependency on its nucleus.

In the same way as *trial,* the lexical constellation of the word *charge* presents a similar structure with respect to the hierarchical levels it has been organised into. However, the LC of *charge* is a more complex one since the secondary level appears to

be more populated. While there is a connection between its primary meanings, either figurative or literal, and the original one, the secondary features which the term acquires bear almost no relationship with the nuclear meaning of the LC. As a matter of fact, an "attack against the enemy" and "an accusation before a judge" do not apparently seem to have anything in common, although, thanks to the possibilities offered by the LC model, this connection can be traced back by following an inverse path from the secondary meanings towards the LC nucleus.

The different semantic senses of the sub-technical term *charge* are illustrated by the examples below obtained from *BLaRC* and *LACELL:*

1) ... relative to his having taken the sole <u>charge</u> of the Spanish detained Vessels sent into Malta ... (*LACELL*)

2) and 6) Again all absolutely free of <u>charge</u>. We don't charge you a penny. No license fee. (*LACELL*)

3) Bomb attack on Canary Wharf, Isle of Dogs, foiled after van laden with ton of explosives spotted by two security guards. Detonator exploded but failed to set off main <u>charge</u>. (*LACELL*)

4) I cried a lot coming out of my teens, says Charlotte, now 22, because I realised that I no longer had an excuse to play out the role of mother's beloved <u>charge</u>. (*LACELL*)

5) Accordingly, an acquittal on a criminal <u>charge</u> of assault based on an assertion by the defendant of the need for self-defence does not mean that the defendant ... (*BLaRC*)

7) Alter the main battery leads to use the front battery to start and run the vehicle and wire the battery under the seat to a split <u>charge</u> relay to take your auxiliary power for a caravan. (*LACELL)*

8) If a charging unit is suddenly confronted by emerging Fanatics, leaving Fanatics in front of the chargers and between them and their target unit, then the chargers can either remain halted or complete their <u>charge</u>. (*LACELL*).

4.4.3. *Battery*

Unlike *trial* or *charge,* the original meaning of this word appears to be the closest one to its legal sense and the one which all other meanings stem from. The *OED* (2002) documents its first use in the 1530s when it was employed to denote "the act of repeatedly beating something or somebody", entering English through Old French (12th c.) *baterie* from the Latin *battuere,* "to beat". The word soon turned into "a unit of artillery" (1550s) not being used to refer to the source of electric energy until the 18th century, for obvious reasons.

Battery is not as frequent as *charge* or *trial* although it could also be considered a highly general word. It appears amongst the first 2,000 word families of *BNC* yet it cannot be found amongst the most frequent 3,000 word families of West's (1953) *GSL* or Coxhead's (2000) *AWL,* as opposed to the other two words examined in this section.

The normalised frequency value of *battery* in the general corpus, 13.35, confirms this fact. Even so, its frequency in the specialised corpus is almost twice as high as it is in the general one, that is, 22.59. Nevertheless, it occurs much less often than *trial* or *charge,* as shown above. This is basically due to the fact that the last two terms are more general within the field of legal English denoting two concepts which are present in most judicial decisions where trial sessions are recorded and charges are always presented and argued.

Judging by its immediate collocates in the general corpus, the legal meaning of *battery* appears to be the most frequent one in the general field, similarly to *trial.* Applying mutual information to identify its collocates in *LACELL* above the >30 frequency threshold established, it only generates two, *assault* (in the first position) and *operated,* pointing at the fact that its specialised sense might be considerably widespread in general English. As for *BLaRC,* its main collocates are *claim, assault* and

damages, only attracting 2 general words (*laptop* and *based*) amongst a list of 11 collocates, which stand at the bottom of the list. The following concordances obtained from *LACELL*, serve as examples of the specialised use of *battery* in a general corpus:

> - They may use gross insults to intimidate, and endearments to redefine what they have done as consensual sex or love instead of rape, assault and battery.
> - In simple terms, battery is the application of unlawful force to another person, whereas assault consists of causing another person to apprehend or expect the application of unlawful force.

In spite of its lower frequency in both corpora, owing to the number and frequency of its collocates and co-collocates in the lexical networks calculated for this term, the sub-technicality coefficient of *battery* is slightly higher than *trial* and *charge,* that is, 5.615. It is 1.1 points above the average of this value for the whole list of terms examined, which situates it closer to the most specialised set of terms studied in section 4.3. The fact that the different meanings of the word stem from the technical one, as illustrated by figure 20, might bear a certain relationship with this data.

The main semantic features offered by the *OED* (2002) are as follows:

> **1)** The act of beating or pounding repeatedly.
> **2)** The unlawful and unwanted touching or striking of one person by another with the intention of bringing about a harmful or offensive contact: *battery and assault.*
> **3)** A set of guns or other heavy artillery, as on a warship.
> **4)** An army artillery unit, corresponding to a company in the infantry.
> **5)** An array of similar things intended for use together: *took a battery of achievement tests.*
> **6)** An impressive body or group*: a battery of political supporters*.
> **7)** Two or more connected cells that produce a direct current by converting chemical energy to electrical energy.
> Electric device which manages to store and provide electric energy make other electric devices run autonomously.
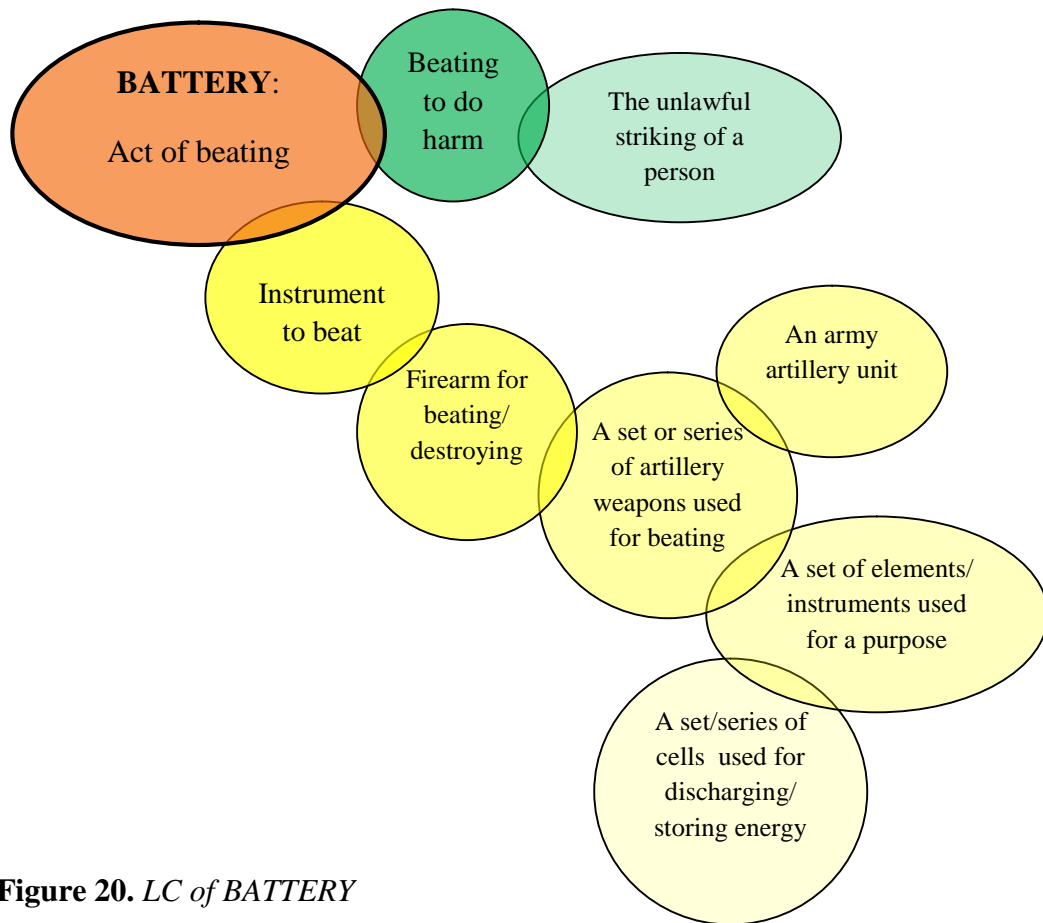
Figure 20. *LC of BATTERY*

Concerning the hierarchy of semantic dependencies portrayed by the LC of *battery,* as indicated by figure 20, it is far more complex than the previous two because of the greater number of levels this LC displays. While the basic meaning of *battery,* "act of beating", is closely linked to its legal usage, trying to associate it with "a set/series of cells used for discharging/storing energy", which ranks fifth as regards its semantic dependency on the rest of constituents of the LC, would not be feasible unless an analysis of this kind was performed. Let us examine in greater detail how this complex structure works.

In contrast to *trial* or *charge,* the base meaning of *battery* only spreads out into two primary senses, the first one, with clearly negative connotations ("beating to do harm"), generates its legal sense, that of unlawfully striking somebody, at a lower

semantic level. The connection between this secondary level and the nucleus of the LC appears to be clear only adding a negative connotation to it through its first level and the sense of *unlawfulness* through the second one.

Regarding the second branch which the base meaning of *battery* splits into, "instrument to beat", it acquires new semantic features as the LC expands, thus generating a greater number of dependencies within its structure. At a third level, *battery* could be understood as a "firearm for beating or destroying" which, in turn, acquires a new feature: "a set or a series of artillery weapons used for beating". This last meaning also evolves into a new one which stems directly from it since the word *battery* is also employed to refer to "an army artillery unit", that is, the group of soldiers who can use artillery weapons to attack. This sense grows parallel to the last two which depart from the military use of the term being employed with a general reference. The concordance below illustrates this parallel semantic sense:

- The whole camp is covered with monuments showing where each <u>battery</u> and other divisions fought ... (*LACELL*).

Finally, the idea of "set or group of elements" connects the fifth and sixth levels of the LC, which have a clearly general character, unlike the upper levels analysed above. The example "a battery of questions", which could be interpreted as a numerous amount of questions formulated with the aim of eliciting information about something, illustrates the fifth semantic level of the structure, that is, "a set of elements/instruments used for a purpose". Stemming from this last sense, *battery* could also be defined as "a set of cells used for discharging or storing energy" once more insisting on the idea of "repeated series of similar elements".

Owing to the greater complexity of this network, which entails the existence of a larger amount of semantic features progressively added to the nucleus of the LC, the hierarchy of meanings is more complex and therefore those senses standing at a lower level present almost no connection (not even figuratively) with the basic or original meaning at the centre of the constellation or with other senses deriving from other primary branches. Probably, as stated by Rea and Sánchez (2010), if the LCs of other words related to *battery* were examined using this model, some of their lower ranking senses might overlap with those standing at the bottom of the semantic hierarchy depicted by the LC of this word, resembling the way star systems connect to each other to form constellations or the structure of neural networks, as already explained above.

To sum up, the visualisation of the LCs allows us to understand more clearly how sub-technical terms, by partially combining their original semantic features with new ones (like the braches of a tree which stem from its trunk), are generated through the acquisition of new specialised meanings to denote new concepts related to already existing ones, hence their shared character. Consequently, the application of this model to such a relevant characteristic of the legal lexicon as its shared nature, facilitates greatly its understanding by offering a multi-dimensional picture of this complex process that could not probably be accounted for otherwise.

4.5. CONCLUSION

This chapter has presented an analysis of sub-technical vocabulary from both a quantitative and a semantic perspective. Quantifying such a phenomenon as this is a complex task due to the fact that this type of words are shared both by the general and

specialised fields and therefore the statistical data associated to them might be misleading when attempting to determine their level of specialisation.

This is the reason why the algorithm *Sub-Tech* was proposed in order to establish a word's sub-technicality level by resorting to its context both in *BLaRC*, the legal corpus, and *LACELL*, the general one. After extracting the specialised terms in the legal corpus applying the most efficient ATR methods tested, *Terminus 2.0* (Nazar and Cabré, 2012) and *TermoStat* (Drouin, 2003), a sample of those terms were selected for varied reasons. They could either be shared by both areas of English without changing their meaning (denoting a legal concept in both cases) or they could become technical and acquire a new specialised meaning when in contact with a legal context.

Once the word selection was made, William's (2001) lexical network model was applied which provided data on the number and frequency of collocates and co-collocates in each word's general and specialised networks. Such data were processed applying a formula whereby the average frequency of the elements in the specialised networks was calculated and compared with the same parameter in the general ones. The result of such comparison led to the ranking of sub-technical words along a continuum of specialisation.

There seems to be no correlation between the results obtained and the semantic categories described in section 4.2. Nevertheless, it appears that those words whose specialised meaning differs greatly from its general one tend to be less distant from the specialised set of words used to test the method than those whose meaning is closer to its general sense. However, a wider sample of words would be necessary in order to reach more definite conclusions in this respect.

One of the advantages of this ranking method is that it is not domain-dependent. On the contrary, it can be applied to any specialised field as long as the comparison with

a general corpus is feasible. Its limitations are basically related to the methods employed to obtain the data. On the one hand, the statistical data associated to each word must be significant enough for such measures as mutual information to work properly, therefore, low frequency words could not be studied applying this technique. On the other hand, William's method produces such a vast amount of data that the networks usually become unmanageable thus requiring the researcher to establish a frequency threshold and to limit the network levels analysed for practical reasons.

All the same, *Sub-Tech* could be regarded as a first attempt towards an objective characterisation of such ambiguous lexical elements as these, something which, to the best of our knowledge, has not been accomplished to date.

The final part of this chapter has been devoted to a semantic description of some of the sub-technical words examined in section 4.4. Implementing Cantos and Sánchez's (2001) lexical constellation model has definitely contributed to a better understanding of the process undergone by sub-technical vocabulary towards specialisation. The LCs of *battery, trial* and *charge* visually illustrate how new semantic senses are added to their base or original meaning presenting a multi-dimensional picture of the dependencies existing amongst the distinguishing semantic features of these shared terms as they evolve towards the acquisition of purely technical meanings.

CHAPTER 5
CONCLUSION, LIMITATIONS OF THIS THESIS AND FURTHER RESEARCH

CHAPTER 5

Overall, the major objectives accomplished by this thesis have been, firstly, to identify the specialised vocabulary in *BLaRC,* an *ad hoc* legal corpus of judicial decisions of 8.85 million words, which is described and justified in detail in chapter 2. In order to do so, ten different ATR methods have been applied to a 2.6 million word corpus, *UKSCC,* extracted from the main one to facilitate their implementation and validation process.

Chapter 3 has therefore been devoted to the evaluation of such ATR methods as regards the precision levels achieved in term identification by each of them. Average precision was calculated through the automatic comparison of the lists of CTs produced by each method with a gold standard, that is, an electronic legal glossary of 10,088 entries. Cumulative precision was also measured following the same procedure so as to observe and compare the way it evolved as the number of identified terms augmented. As a result, *Terminus 2.0* (Nazar & Cabré, 2012) and *TermoStat* (2003), the best performing techniques, were selected with the aim of implementing them on *BLaRC*. After doing so, the validated lists of both single and multi-word legal terms extracted from it have been offered in section 3.2.4. Chapter 3 ends with the proposal of some activities aimed at illustrating the varied applications and uses of specialised corpora and vocabulary lists.

Owing to the relevance of sub-technical vocabulary as a major component of the legal lexicon, a quantitative method has been proposed in chapter 4 to measure its degree of specialisation based on the context of usage of this type of words. William's (2001) lexical network model was applied to a set of general, highly specialised and sub-technical words in order to observe and compare the number and frequency of their

collocates and co-collocates both in *BLaRC,* the specialised corpus, and *LACELL,* the general one. The observation of the data obtained led to the formulation of the algorithm *Sub-Tech* allowing to place the words analysed along a continuum of specialisation depending on the data obtained after the implementation of Williams' model.

Finally, with the purpose of describing sub-technical vocabulary from a semantic perspective, Cantos and Sánchez's (2001) lexical constellation model has been applied to analyse the semantic features of the shared terms *trial, charge* and *battery* resulting into a much clearer picture of the process undergone by this type of words from general usage to specialisation. The application of this model in combination with the quantitative method described above may be regarded as a first step towards a better understanding of a lexical phenomenon which has not been explored in depth to date.

In spite of the above, there are certain limitations in this work that could not be avoided and which should be taken into account for further research. To begin with, including other legal genres in *BLaRC* might be desirable so as to allow the conclusions based on it to be more representative of the whole legal variety and not only restricted to a single genre, although law reports, due to their role within common law legal systems, could be deemed fundamental for this sub-language.

In addition, it would be interesting to establish a correlation between the data provided through the implementation of the algorithm *Sub-Tech* and the semantic features of sub-technical terms included in the taxonomy offered in section 4.2. In order to do that, it may be necessary to increase the number of words analysed so that the data obtained could be more representative of this type of vocabulary.

It might also be argued that, by removing the threshold applied to obtain the lexical networks used in chapter 4, they could include all the elements attracted by the node and its collocates, co-collocates and so forth, and thus provide a fuller picture of their context of usage and more accurate data to base our conclusions on. Nevertheless, they would become unmanageable requiring the automatisation of the method to obtain and process such a vast amount of data, which has not taken place as yet.

Lastly, regarding the lexical constellation model, it might be an interesting initiative to try and analyse other terms or sub-technical terms that may be associated in any way to the ones examined in section 4.4. (i.e. their most relevant collocates) to attempt to demonstrate how their constellations might overlap with each other thus corroborating the way these networks interweave on a higher semantic level.

REFERENCES

Ahmad, K., Davies, A., Fulford, H., Rogers, M. (1994). "What is a term? The semi-automatic extraction of terms from text". In Snell-Hornby, M., Pöchhacker, F., Kaindl, K. (eds.) *Translation Studies: An Interdiscipline.* Amsterdam: John Benjamins, 267-278.

Ahmad, K., Gillam, L., Tostevin, L. (2000). "Weirdness Indexing for Logical Document Extrapolation and Retrieval (WILDER)". In *Proceedings of the Eighth Text Retrieval Conference (TREC-8).* National Institute of Standards and Technology: U.S. Department of Commerce.

Alcaraz Varó, E.

----------- (1994). *El inglés jurídico: textos y documentos.* Madrid: Ariel Derecho.

----------- (2000). *El inglés profesional y académico*. Madrid: Alianza Editorial.

Alcaraz Varó, E. Hughes, B. (1993). *Diccionario de términos jurídicos.* Barcelona: Ariel.

Almela, A. (2008). *Evaluating Multiword Automatic Term Recognition Techniques on a Veterinary Medicine Corpus.* Tesis fin de Máster. Universidad de Murcia.

Almela, M. (2006). *From words to lexical units: A corpus-driven account of collocation and Idiomatic Patterning in English and English-Spanish.* Frankfurt: Peter Lang.

Almela, M., Sánchez, A., Cantos, P. (2004). "Lexico-semantic mappping of meanings in English and Spanish. A model of analysis". In J.M. Bravo (ed.). *Aspects of Translation*. Valladolid: Universidad de Valladolid.

Alonso, M. (1994)."Hacia una definición del concepto de colocación: de J.R. Firth a Mel'cuk". *Revista de lexicografía,* I: 9-28.

Alonso, M., Millon, C., Williams, G. (2011). "Collocational Networks and their Applications to an E-Advanced Learner"s Dictionary of verbs in Science (DicSci)". In *Proceedings of eLex 2011.* Bled, Slovenia. November, 2011.

Allwright, R. L. (1981). "What do we want teaching materials for?". *ELT Journal*, 36 (1): 5–18.

Ananiadou, S.

----------- (1988). *A Methodology for Automatic Term Recognition.* PhD Thesis, University of Manchester Institute of Science and Technology: United Kingdom.

------------- (1994). "A methodology for automatic term recognition". In *COLING. Proceedings of the 15th International Conference on Computational Linguistics*, 1034-1038.

Arson, A., Lang, F. 2010. 'An overview of MetaMap: historical perspective and recent advances. *Journal of American Medical Informatics Association,* 17: 229-236.

Aston, G. (1997). "Enriching the learning environment: corpora in ELT". In Wichmann, A. et al. (eds.). *Teaching and language corpora*. London: Longman.

Baker, M. (1988). "Sub-technical vocabulary and the ESP teacher: An analysis of some rhetorical items in medical journal articles". *Reading in a Foreign Language,* 4 (2): 91-105.

Barber, C.L. (1962). "Some measurable characteristics of modern scientific prose". In Behre, F. (ed.). *Gothenburg Studies in English: Vol. 14. Contributions to English Syntax and Philology.* Gothenburg: University of Gothenburg.

Barrón-Cedeño, A., Sierra, G. E., Drouin, P., Ananiadou, S. (2009). "An Improved Automatic Term Recognition Method for Spanish". In Gelbukh, A. (ed.).

Proceedings of the 10th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing 2009). New York: Springer.

Bauer, L., Nation, P. (1993). "Word Families". *International Journal of Lexicography, 6* (4): 253-279.

Bernier-Colborne, G. (2012). "Defining a Gold Standard for the Evaluation of Term Extractors". In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC"12).* Instambul, Turkey. May 2012. URL: http://www.lrec-conf.org/proceedings/lrec2012/index.html

Bhatia, V.

----------- (1993). *Analysing Genre: Language Use in Professional Settings*. London: Longman.

----------- (2004). "Applied genre analysis: a multi-perspective model". *Iberica 4*: 4-19.

Biber, D.

----------- (1988). *Variation across Speech and Writing*. Cambridge: Cambridge University Press.

----------- (1993). "Representativeness in corpus design". In *Literary and Linguistic Computing,* 8 (4): 243-57.

Biber, D., Conrad, S., Reppen, R. (1998). *Corpus Linguistics. Investigating Language Structure and Use.* Cambridge: Cambridge University Press.

Borja Albí, A. (2000). *El texto jurídico en inglés y su traducción.* Barcelona: Ariel.

Boulton, A. (2011). "Data Driven Learning: the Perpetual Enigma". In Roszkowski, S., Lewandowska-Tomaszczyk, B. (eds.). *Explorations across Languages and Corpora*. Frankfurt: Peter Lang.

Bourigault, D. (1992). "Surface grammatical analysis for the extraction of terminological noun phrases". In *Proceedings of the 5th International Conference on Computational Linguistics.* Nantes, France. August, 1992.

Breeze, R. (2011). "Disciplinary values in legal discourse: a corpus study". *Ibérica,* 21: 93-116.

British National Corpus, version 3 (*BNC* XML Edition). (2007). Distributed by Oxford University Computing Services on behalf of the *BNC* Consortium. URL: http://www.natcorp.ox.ac.uk

Brodine, R. (2001). "Integrating Corpus Work into an Academic Reading Course". In Aston, G. (ed.). *Learning with Corpora.* Bologna: Athelstan.

Brown, G.D., Rice, S. (2007). *Professional English in Use: Law.* Cambridge: Cambridge University Press.

Cabré, M.T. (1993). *La terminología. Teoría, metodología, aplicaciones.* Barcelona: Antártida/Empúries.

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

Callanan, H., Edwards, L. (2010). *Absolute Legal English.* London: Delta.

Cantos, P. (2013). *Statistical Methods in Language and Linguistic Research.* Sheffield: Equinox Publishing.

Cantos, P., Sánchez, A. (2001). "Lexical Constellations: What Collocates Fail to Tell". *International Journal of Corpus Linguistics,* 6 (2): 199-228.

Cheng, Winnie. (2011). *Exploring corpus linguistics: Language in action.* Abingdon: Routledge.

Chujo, K., Utiyama, M., Nakamura, T., Oghigian, K. (2010). "Evaluating statistically-extracted domain-specific word lists". In *Corpus, ICT, and Language Education.* Weir, G., Ishikawa, S. (eds.). Glasgow: University of Strathclyde Publishing.

Chung, T. M.

---------- (2003a). "A corpus comparison approach for terminology extraction". *Terminology,* 9 (2): 221-246.

---------- (2003b). *Identifying Technical Vocabulary.* PhD thesis. Victoria University of Wellington.

Chung, T.M., Nation, P. (2003). "Technical Vocabulary in Specialised Texts". *Reading in a Foreign Language,* 15 (2): 103-116.

Church, K.W., Hanks, P. (1990). "Word association norms, mutual information, and lexicography". *Computational Linguistics,* 16 (1): 22-29.

Church, K.W., Gale, W. (1995). "Inverse Document Frequency (IDF): A measure of Deviations from Poisson". *Proceedings of the Third Workshop on Very Large Corpora.* Cambridge: Massachusetts Institute of Technology.

Cohen, A., Glasman, H., Rosenbaum-Cohen, P.R., Ferrara, J., Fine, J. (1988). "Reading English for Specialised Purposes: discourse analysis and the use of student informants". In Carrell, P., Daine, J., Eskey, D.E. (eds.). *Interactive Approaches to Second Language Reading.* Cambridge: Cambridge University Press.

Corpas Pastor, G. (2003). *Recursos documentales y tecnológicos para la traducción del discurso jurídico (español, alemán, inglés, italiano, árabe).* Granada: Comares.

Corpas Pastor, G., Seghiri Dominguez, M. (2010). *El concepto de representatividad en lingüística de corpus: aproximaciones teóricas y consecuencias para la traducción.* Málaga: Servicio de Publicaciones de la Universidad de Málaga.

Cotter, M.J. (2006). "Teaching terms: a corpus-based approach to terminology in ESP classes". In *Proceedings of the 5th International AELFE conference.* Zaragoza, Spain. September, 2006.

Cowan, J.R. (1974). "Lexical and syntactic research for the design of EFL". *TESOL Quarterly,* 8: 389-99.

Coxhead, A. (2000). "A New Academic Word List". *TESOL, Quarterly,* 34 (2): 213-238.

Coxhead, A., Nation, P. (2001). "The Specialised Vocabulary of English for Academic Purposes". Flowerdew, J., Peacock, M. *Research Perspectives on English for Academic Purposes*. Cambridge: Cambridge University Press.

Cruse, A. (2004). *Meaning in Language: AnIntroduction to Semantics and Pragmatics*. Oxford: Oxford University Press

Cruz Martínez, M.S. (2002). "La colocación léxica y gramatical en el proceso penal inglés". *Ibérica,* 4: 129-143.

Crystal, D., Davy, D. (1969). *Investigating English Style.* London: Longman.

Daille, B. (1996). "Study and implementation of combined techniques for automatic extraction of terminology". In Klavans, J.L., Resnik, P. (eds.). *The Balancing act: Combining symbolic and statistical approaches to language.* Cambridge, MA: MIT Press.

Danet, B. (1980). "Language in the Legal Process". *Law and Society Review,* 14 (3): 445-564.

David, S., Plante, P. (1990). *Termino* 1.0. Research Report of Centre d'Analyse de Textes par Ordinateur. Université du Québec, Montréal.

Dresher, R. (1934). "Training in mathematics vocabulary". *Educational Research Bulletin,* 13: 201-4.

Drouin, P.

---------- (2003). "Term extraction using non-technical corpora as a point of leverage". *Terminology,* 9 (1): 99-117.

---------- (2004). "Detection of Domain Specific Terminology Using Corpora Comparison". *Proceedings of the fourth international Conference on Language Resources and Evaluation (LREC).* Lisbon, Portugal. May, 2004.

Dudley-Evans, T., St John, M. (1998). *Developments in English for Specific Purposes.* Cambridge: Cambridge University Press.

Dunning, T. (1993). "Accurate Methods for the Statistics of Surprise and Coincidence". *Computational Linguistics*, 19 (1): 61-74.

Farrell, P. (1990). *Vocabulary in ESL: A Lexical Analysis of the English of Electronics and a Study of Semi-technical Vocabulary.* Dublin: Centre for Language and Communication Studies.

Fahmi, I., Bouma, G., Van der Plas, L. (2007). 'Improving statistical method using known terms for automatic term extraction'. In *Computational Linguistics in the Netherlands. CLIN,* 17: 1-8.

Flowerdale, L. (2004). "The argument for using English specialised corpora to understand academic and professional language." In U. Connor and T. Upton, (eds.). *Discourse In The Professions: Perspectives From Corpus Linguistics.* Amsterdam/Philadelphia: John Benjamins.

Flowerdew, L.

------------ (2001). "Concordancing as tool in course design". In Ghadessi, M., Henry, A., Roseberry, R. (eds.) *Small corpus studies and ELT: Theory and practice.* Amsterdam: John Benjamins.

------------ (2009). "Applying corpus linguistics to pedagogy: A critical evaluation". *International Journal of Corpus Linguistics,* 14 (3): 393-417.

Frantzi, K.T., Ananiadou, S.

------------ (1996). "Extracting nested collocations". In *Proceedings of the 16th Conference on Computational Linguistics*. Copenhaguen, Denmark. August, 1996.

------------ (1999). "The c/nc value domain independent method for multi-word term extraction". *Journal of Natural Language Processing,* 3 (2): 115-127.

Fraser, S. (2007). "Providing ESP Learners with the Vocabulary They Need: Corpora and the Creation of Specialized Word Lists". *Hiroshima Studies in Language and Language Education*, 10: 127-143.

Fuertes-Oliviera, P.A. (2008). "Pedagogical Application of Specialized Corpora in ESP Teaching. the case of the UVaSTECorpus". *Scripta Manent,* 3 (2): 68-81.

Gabrielatos, C., Marchi, A. (2011). "*Keyness : Matching metrics to definitions"*. In *Theoretical-methodological challenges in corpus approaches to discourse studies and some ways of addressing them*. Portsmouth, United Kingdom. August, 2011 (unpublished).

URL: http://eprints.lancs.ac.uk/51449/4/Gabrielatos_Marchi_Keyness.pdf

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

Gries, S., Wulff, S. (eds.) (2010). *Corpus-linguistics applications. Current studies, new directions.* Amsterdam/New York: Rodopi.

Harris, S. (1997). "Procedural Vocabulary in Law Case Reports". *English for Specific Purposes,* 16 (4): 289-308.

Harwood, N. (2005). "What do we want EAP teaching materials for?". *Journal of English for Academic Purposes,* 4: 149-161.

Heatley, A., Nation, P. (1996). *Range* (computer software)*.* Wellington, New Zealand: Victoria University of Wellington.

URL: http://www.victoria.ac.nz/lals/about/staff/paul-nation

Heaps, H. S. (1978). *Information Retrieval: Computational and Theoretical Aspects*. New York: Academic Press.

Hoffmann, S., Evert, S., Smith, N., Lee, D., Berglund-Prytz, Y. (2008). *Corpus Linguistics with BNCweb: A Practical Guide*. Frankfurt am Main: Peter Lang.

Hsu, W. (2011). "The vocabulary thresholds of business textbooks and business research articles for EFL learners". *English for Specific Purposes,* 30: 247-257.

Hutchison, T., Waters, A. (1998). *English for Specific Purposes*. Cambridge University Press.

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

Hyland, K. & P. Tse (2007). "Is there an 'Academic Vocabulary'?". *TESOL Quarterly,* 41 (2): 235-253.

Igor, J. (The Lord Chief of Justice of England and Wales) (2009). *Introductory Speech at the ICLR Launch of Always cited in Preference.* The Old Hall, Lincoln's Inn, London. October 21st.

Jackson, B. (1985). *Semiotics and Legal Theory.* London: RKP.

Jacquemin, C. (2001). *Spotting and discovering terms through NLP.* Massachusetts: MIT Press.

Johns, T. (1991). "Should you be persuaded: Two examples of data-driven learning". In Johns, T., Kings, P. (eds.). *Classroom Concordancing English Language Research Journal,* 4. Birmingham: University of Birmingham.

Joos, M. (1961). *The Five Clocks: A Linguistic Excursion into the Five Styles of English Usage.* New York: Hancourt, Brace & Court.

Justeson, J.S., Katz, S.M. (1995). "Technical terminology: some linguistic properties and an algorithm for identification in text". In *Natural Language Engineering*, 1: 9-27.

Kaltenböck, G., Mehlmauer-Larcher, B. (2005). "Computer corpora and the language classroom: forced marriage or perfect match? On the potential and limitations of computer corpora in language teaching". *ReCall,* 17 (1): 65–84.

Kennedy, G. (1998). *An introduction to corpus linguistics*. New York: Longman.

Kennedy, G. y Bolitho, R. (1984). *English for specific purposes*. London: Mcmillan.

Kilgariff, A. et *al.* (2003). *Sketch Engine.* Online computer software. Lexical computing Ltd.

Kilgarriff, A. (2006). "Word Senses". In Aguirre, E., Edmonds, P. (eds.). *Word Sense Disambiguation. Algorithms and Applications.* Berlin: Springer Verlag.

Kit, C., Liu, X. (2008). "Measuring mono-word termhood by rank difference via corpus comparison". *Terminology*, 14 (2): 204-229.

Krois-Linder, A., Firth, M. (2008). *Introduction to International Legal English : A course for classroom or self-study use*. Cambridge: Cambridge University Press.

Kurzon, D. (1984). "Themes, Hyperthemes and the Discourse Structure of English Legal Texts". *Text,* 4: 1-3.

Lam, J.K. (2001). *A study of semi-technical vocabulary in Computer Science texts, with special reference to ESP teaching and lexicography.* PhD thesis. University of Exeter.

URL: http://hdl.handle.net/1783.1/1056

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.

Leech, G.

----------- (1991). "The state of the art in corpus linguistics." In Aijmer, K., Altenberg, B. (eds.) *English corpus linguistics: Studies in Honour of Jan Svartvik*. London: Longman.

----------- (1997). "Teaching and language corpora: A convergence". In Wichmann, A., Fligelstone, S., McEnery, T., Knowles, G. (eds.). *Teaching and language corpora.* London: Longman.

Loginova, E., Gojun, A., Blancafort, E., Guegan, M., Gornostay, T., Heid, U. "Reference Lists for the Evaluation of Term Extraction Tools". In *TKE 2012: Terminology and Knowledge Engineering*, 19th-22nd June, Madrid. URL: http://www.ttc-project.eu/images/stories/TTC_TKE_2012.pdf

Maley, Y.

------------ (1987). "The Language of Legislation". *Language and Society,* 16: 25-48.

------------ (1994). "The Language of the Law". In Gibbons, J. (ed). *Language and the Law*. London: Longman.

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., 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. Abril, 2011.

Marín, M.J. and Rea, C. (2012). "Structure and design of the BLRC: a legal corpus of judicial decisions from the UK". *Journal of English Studies*, 10. La Rioja: Servicio de Publicaciones de la Universidad de La Rioja. URL: http://www.unirioja.es/servicios/sp/ej/jes/jes10.shtml

Marín, M.J., Rea, C. (2012). "How relevant are Latin wordforms and clusters in legal English? A corpus-based study on the representativeness and specificity of such elements in UKSCC: an ad hoc legal corpus". *ES. Revista de Filología Inglesa, 33*. Valladolid: Servicio de Publicaciones de la Universidad de Valladolid: 161-182.

Marín, M.J. (forthcoming, 2014). "Evaluation of five single-word term recognition methods on a legal corpus". *Corpora,* 9 (1). Endinburgh: Endinburgh University Press.

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

Mellinkoff, D. (1963). *The Language of the Law.* Boston: Little, Brown & Co.

Meyer, Ch. (2002). *English corpus linguistics: An introduction.* Cambridge: Cambridge University Press.

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

McEnery, T., Xiao, R., Tono, Y. (2006). *Corpus-based language studies: an advanced resource book.* Routledge Applied Linguistics: New York.

McEnery, T., Xiao, R. (2011). "What corpora can offer in language teaching and learning". In Hinkel, E. (ed.), *Handbook of Research in Second Language Teaching and Learning.* London: Routledge

Mondary, T., Nazarenko, A., Zargayouna, H., Berreaux, S. (2012). "The Quaero Evaluation Initiative on Term Extraction". In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12).* Instambul, Turkey. May, 2012. URL: http://www.lrec-conf.org/proceedings/lrec2012/index.html

Murphy W.T., Rawlings, R.W. (1981). "After the Ancient Regime: The Writing of Judgments in the House of Lords 1979/1980". *The Modern Law Review,* 44 (6): 617-657. URL: http://www.jstor.org/stable/1095030

Nakagawa, H., Mori, T. (2002). "A simple but powerful automatic term extraction method". In *COLING-02 on COMPUTERM. Proceedings of the Second International Workshop on Computational Terminology.* Taipei, Taiwan. August, 2002.

Nation, P.

------------ (1990). *Teaching and Learning Vocabulary.* Newbury House: New York.

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

Nation, P., Waring R. (1997). "Vocabulary Size, Text Coverage and Word Lists". In Schmitt, N., McCarthy, M. (eds.). *Vocabulary: Description, Acquisition and Pedagogy*. Cambridge: Cambridge University Press.

Navigli, R., Velardi, P. (2002). "Semantic Interpretation of Terminological Strings". In *Proceedings of the 6th International Conference on Terminology and Knowledge Engineering (TKE2002).* Vandoeuvre-lès-Nancy, France. August, 2002.

Nazar, R., Cabré, M.T. (2012). "Supervised Learning Algorithms Applied to Terminology Extraction". In Aguado de Cea, G., Suárez-Figueroa, M.C., García-Castro, R., Montiel-Ponsoda, E. (eds.). *Proceedings of the 10th Terminology and Knowledge Engineering Conference* (TKE 2012). Ontology Engineering Group, Association for Terminology and Knowledge Transfer. Madrid, Spain. June, 2012.

Nelson, M. (2000). *A corpus-based study of the lexis of business English and business English teaching materials.* PhD Thesis. University of Manchester.

Orts Llopis, M.A.

 ------------ (2006). *Aproximación al discurso jurídico en inglés : las pólizas de seguro marítimo de Lloyd's.* Madrid: Edisofer.

------------ (2009). "Legal genres in English and Spanish: some attempts of analysis". *Iberica,* 18: 109-130.

Pamies Beltrán, A., Pazos Bretaña, J.M. (2003). "Acceso automatizado a fraseologismos y colocaciones en corpus no etiquetado". *Language Design,* 5: 39-50. URL: http://elies.rediris.es/Language_Design/LD5/indice_vol5.html

Park, Y., Byrd, R.J., Boguraev, B. (2002). "Automatic Glossary Extraction: Beyond Terminology Association". In *Proceedings of COLING "02 19th International Conference on Computational Linguistics.* Taipei, Taiwan. August, 2002.

Partington, A. (1998). *Patterns and Meanings. Using Corpora for English Language Research and Teaching*. Amsterdam: John Benjamins.

Panzienza, M.T., Pennacchiotti, M., Zanzotto, F.M. (2005). 'Terminology extraction: An Analysis of Linguistic and Statistical Approaches'. *Studies in Fuziness and Soft Computing,* 185: 225-279.

Pearson, J. (1998). *Terms in Context*. Amsterdam: John Benjamins.

Perell, P.M. (1987). "Stare decisis and techniques of legal reasoning and legal argument". URL: http://www.legalresearch.org/docs/perell.html

Pérez, P., Cantos, P. (2004) "Some lessons students learn: Self-discovery and corpora". In Aston, G., Bernardini, S., Dominic, S. (eds). *Corpora and Language Learners*. Amsterdam-Philadelphia: John Benjamins.

Räisänen, C., Fortanet, I. (eds.). (2008). *ESP in European Higher Education: Integrating language and content*. Amsterdam: John Benjamins.

Rea, C.

----------- (2008). *El inglés de las telecomunicaciones: estudio léxico basado en un corpus específico.* PhD Thesis. Universidad de Murcia.

URL: http://www.tdx.cat/handle/10803/10819

---------- (2010). "Getting on with Corpus Compilation: from Theory to Practice". In *ESP World,* 1 (27).

URL: http://www.esp-world.info/articles_27/camino%20rea.pdf

Rea, C., Sánchez, A. (2010). "Building new meanings in technical English from the perspective of the lexical constellation model". *Iberica,* 20: 107-126.

Read, J. (2007). Second Language Vocabulary Assessment: Current Practices and New Directions. *International Journal of English Studies*, 7 (2): 105-26.

Rice, S., Brown, G. (2007). *Professional English in use: Law.* Cambridge: Cambridge University Press.

Robertson, S. (2004). "Understanding Inverse Document Frequency: On theoretical Arguments for IDF". In *Journal of Documentation,* 60: 503-520.

Rodgers, O., Chambers, A., Le Baron-Earle, F. (2011). "Corpora in the LSP classroom: A learner-centered corpus of French for biotechnologists". In *International Journal of Corpus Linguistics,* 16 (3): 391-411.

Römer, U. (2008). "Corpora and Language Teaching". In Lüdeling, A., Kyto, M. (eds.). *Corpus Linguistics. An International Handbook.* Volume I. Berlin: Mouton de Gruyter: 112-130.

Rossini, R., Tamburini, F., Martelli, E. (2001). "Words from the Bononia Legal Corpus". *International Journal of Corpus Linguistics,* 6: 13-34.

Saint Dahl, H. (1999). *Dahl's Law Dictionary. Diccionario Jurídico Dahl.* New York: William S. Hein & Co., Inc.

Sánchez, A. (2000), 'Language Teaching before and after 'Digitilized Corpora'. Three Main Issues'. *Cuadernos de Filología Inglesa*, 9.1. Murcia: Servicio de Publicaciones, Universidad de Murcia.

Sánchez, A., Cantos, P., Sarmiento R., Simón, J. (1995). *Cumbre. Corpus lingüístico del español contemporáneo. Fundamentos, metodología y análisis.* Madrid: SGEL.

Sánchez A., Cantos, P. (1997). "Predictability of Word Forms *(*Types*)*, and Lemmas in Linguistic Corpora". *International Journal of Corpus* Linguistics*,* 2 (2): 251-272.

Sánchez, A., Cantos, P., Almela, M. (2010). "In search of a new model for WSD: a multi-layered Lexical Constellation Model". In Sánchez, A., Almela, M. (eds.). *A Mosaic of Corpus Linguistics: Selected Approaches.* Frankfurt: Peter Lang.

Sclano, F., Velardi, P., (2007). "A Web Application to Learn the Common Terminology of Interest Groups and Research Communities". *Proceedings of the Conference TIA-2007.* Sophia Antipolis, France. October, 2007.

Schmid, H.

---------- (1994). "Probabilistic Part-of-Speech Tagging Using Decision Trees". In *Proceedings of the International Conference on New Methods in Language Processing*, Manchester, UK.

---------- (1995). "Improvements in Part-of-Speech Tagging with an Application to German". In *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland.

Schmitt, N.

---------- (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.

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

Scott, M.

---------- (2008a). *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software.

---------- (2008b). *WordSmith Tools Help*. Liverpool: Lexical Analysis Software.

Scott, M., Tribble, C. (2006). *Textual patterns: Keyword and corpus analysis in language education.* Amsterdam: Benjamins.

Sinclair, J.

\-\-\-\-\-\-\-\-\-\-\- (1984). "Naturalness in Language". In Aarts, J., Meijs, W. (eds.). *Corpus Linguistics. Recent Developments in the Use of Computer Corpora in English Language Research.* Amsterdam: Rodopi.

\-\-\-\-\-\-\-\-\-\-\- (1991). *Corpus, Concordance and Collocation.* Oxford: Oxford University Press.

\-\-\-\-\-\-\-\-\-\-\- (2004). "Intuition and annotation - the discussion continues". In *Advances in corpus linguistics. Papers from the 23rd International Conference on English Language Research on Computerized Corpora (ICAME 23).* Göteborg, Sweden. May 2002.

\-\-\-\-\-\-\-\-\-\-\- (2005). "Corpus and Text: Basic Principles". In Wynne, M. (ed.). *Developing Linguistic Corpora: a Guide to Good Practice.* AHDS Literature, Languages and Linguistics: University of Oxford.
URL: http://ota.ahds.ac.uk/documents/creating/dlc/index.htm

Spasic, I., Ananiadou, S., McNaught, J. & Kumar, A. (2005). "Text mining and ontologies in biomedicine: Making sense of raw text". *Brief Bioinform*, 6 (3): 239-251.

Sparck Jones, K. (1972). "A statistical interpretation of term specificity and its application in retrieval". *Journal of Documentation,* 28: 11-21.

Stubbs, M.

\-\-\-\-\-\-\-\-\- (1993). "British Traditions in Text Analysis: From Firth to Sinclair". In Baker, M., Francis, G., Tognini-Bonelli, E. (eds.) *Text and Technology. In Honour of John Sinclair.* Philadelphia/Amsterdam: John Benjamins.

\-\-\-\-\-\-\-\-\- (1996). *Text and Corpus Analysis*. Oxford: Blackwell.

--------- (2001). *Words and phrases. Corpus studies of lexical semantics*. Oxford: Blackwell.

--------- (2004). "Which words must students know?: A note on "A New Academic Word List" (Coxhead 2000)". URL: http://www.uni-trier.de/uni/fb2/anglistik/Projekte/stubbs

Swales, J.

--------- (1985). *Episodes in ESP*. Hemel: Hempstead Prentice Hall International.

--------- (1990). *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Tetley, W. (2000). "Mixed Jurisdictions: Common Law vs. Civil Law (Codified and Uncodified)". URL: http://www.mcgill.ca/files/maritimelaw/mixedjur.pdf

Thorndike, E.L., Lorge, I. (1944). *The teacher's Word Book of 30,000 Words*. Teachers' College: Columbia University.

Tiersma, P. (1999). *Legal Language.* Chicago: The University of Chicago Press.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work*. Amsterdam: John-Benjamins.

Tribble, C., Jones, G. (1997). *Concordances in the Classroom: A Resource Book for Teachers.* Houston: Athelstan.

Trimble, L. (1985). *English for science & technology: A discourse approach*. Cambridge: Cambrige University Press.

Vivaldi, J., Cabrera-Diego, L.A., Sierra, G., Pozzi, M. (2012). "Using Wikipedia to Validate the Terminology Found in a Corpus of Basic Textbooks". In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12).* Instambul, Turkey. May, 2012.
URL: http://www.lrec-conf.org/proceedings/lrec2012/index.html

Wang, K., Nation, P. (2004). "Word Meaning in Academic English: Homography in the Academic Word List". *Applied Linguistics,* 25 (3): 291-314.

URL: https://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/2004-Wang-AWL.pdf

Ward, J. (2009). "A basic engineering word list for less proficient foundation engineering undergraduates". *English for Specific Purposes,* 28: 170-182.

West, M. (1953). *A General Service List of English Words*. London: Longman.

Widdowson, H. (1998). "Communication and community: The pragmatics of ESP*"*. *English for Specific Purposes* 17 (1): 3-14.

Williams, G. (2001). "Mediating between lexis and texts: collocational networks in specialised corpora". *ASp. la revue du GERAS,* 31-33: 63-76.

Wynne, M. (ed.) (2005). *Developing Linguistic Corpora: a Guide to Good Practice.* ASDS Literature, Languages and Linguistics. Oxford.

Yamamoto, M., Church, K (1998). "Using Suffix Arrays to Compute Term Frequency and Document Frequency for All Substrings in a Corpus". In *Proceedings of the Sixth Workshop on Very Large Corpora*. Montreal, Canada. August, 1998.

Yang, H. (1986). "A New Technique for identifying Scientific/Technical Terms and Describing Science Texts (An Interim Report)". *Literary and Linguistic Computing,* 1 (2): 93-103.

Yang, D., Cantos Gómez, P., Song, M. (2000). "An Algorithm for Predicting the Relationship between Lemmas and Corpus Size." *ETRI Journal* 22 (2): 20-31.

Young-Mi, J. (1995). "Statistical Characteristics of Korean Vocabulary and Its Application". *Lexicographic Study*, 5 (6): 134-163.

Zhang, Z., Iria, J., Brewster, C., Ciravegna, F. (2008). "A Comparative Evaluation of Term Recognition Algorithms". In *Proceedings of the sixth International*

Conference on Language Resources and Evaluation, (LREC 2008). Marrakech,

Morocco. May, 2008.

---------- *Oxford English Dictionary.* (2002) Version 3.0. Electronic Resource. Oxford:

Oxford University Press.

ONLINE RESOURCES

Administrative Justice and Tribunal Council (AJTC)

 URL: http://www.ajtc.gov.uk

British and Irish Legal Information Institute

URL: http://www.bailii.org

Bononia Legal Corpus

URL: http://corpora.dslo.unibo.it/bolc_eng.html

BNC lemmatised lists (Adam Kilgariff)

URL: http://www.kilgarriff.co.uk/BNClists/lemma.num

Cambridge International Corpora

URL:

http://www.cambridge.org/es/elt/catalogue/subject/custom/item3637700/Cambridge-International-Corpus-Cambridge-International-Corpus/?site_locale=es_ES

Cobuild corpora

URL: http://www.collins.co.uk/Corpus/CorpusSearch.aspx

Constitutional Reform Act 2005

URL: http://www.statutelaw.gov.uk/content.aspx?activeTextDocId=1974190

Cor-Tec corpus

URL: *http://www.fflch.usp.br/dlm/comet/projeto.html*

Court sentence finder

URL: http://www.thelawpages.com/

Her Majesty's Court Service

URL: http://www.hmcourts-service.gov.uk

UK Court and Tribunal structure

URL: http://www.hmcourts-service.gov.uk/aboutus/structure/index.htm

HOLJ corpus

URL: http://www.ltg.ed.ac.uk/SUM/CORPUS

International Language and Law Association

URL: http://www.illa.org/

JATE tools (Zang et al.)

URL: http://code.google.com/p/jatetoolkit

JRC-Acquis corpus

URL: http://langtech.jrc.ec.europa.eu/JRC-Acquis.html

Judicial Committee of the Privy Council

URL: http://www.jcpc.gov.uk/

Judiciary of England and Wales. Glossary

URL: http://www.judiciary.gov.uk/glossary

Jurispedia

URL: http://en.jurispedia.org

Justis, online legal library

URL: http://www.justis.com

UK Legislation archives

URL: http://www.legislation.gov.uk/

Lexis Nexis

URL: http://www.lexisnexis.co.uk

Lextutor (Tom Cobb)

URL: http://www.natcorp.ox.ac.uk

List of United Kingdom Tribunals

URL: http://en.wikipedia.org/wiki/List_of_United_Kingdom_tribunals

Mark Davies Corpus.byu

http://corpus.byu.edu

Nolo's Dictionary of Law Terms and Legal Definitions

URL: http://www.nolo.com/dictionary

Northern Ireland Courts and Tribunal Service

URL: http://www.courtsni.gov.uk

Office for National Statistics, UK

URL: http://www.ons.gov.uk/census

Plain English Campaign UK

URL: http://www.plainenglish.co.uk

Scottish Government website

URL: http://www.scotland.gov.uk

Sixth Form Law. Glossary

URL: http://sixthformlaw.info/03_dictionary/index.htm

Sketch Engine (Kilgariff et al., 2004)

URL: http://www.sketchengine.co.uk

Supreme Court of the United Kingdom

URL: http://www.supremecourt.gov.uk

The Scottish Courts Website

URL: http://www.scotcourts.gov.uk

The Incorporated Council of Law Reporting of England and Wales

URL: http://www.lawreports.co.uk/

Law in the United Kingdom

URL: http://en.wikipedia.org/wiki/Law_of_the_United_Kingdom

The English-Chinese glossary of legal terms

URL: http://www.legislation.gov.hk/eng/glossary/homeglos.htm

Termextractor (Sclano and Velardi)

URL: http://lcl.uniroma1.it/termextractor

Terminus 2.0 (Nazar and Cabré)

URL: http://terminus.upf.edu

TermoStat (P. Drouin)

URL: http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger
Tree Tagger (H. Schimd)

URL: http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger
Tribunals, Courts and Enforcement Act 2007

URL: http://www.legislation.gov.uk/ukpga/2007/15/contents

Tribunals Service of the United Kingdom

URL: http://www.tribunals.gov.uk

APPENDICES

BLOCK 1. VALIDATED 200 TOP LEGAL TERMS

TermoStat (Drouin, 2003): SWTs

TERMS	TECHNICALITY	TERMS	TECHNICALITY
SECTION	126.29	TENANT	44.75
V. (VERSUS)	112.55	POSSESSION	44.51
CASE	111.79	PARAS. (PARAGRAPHS)	44.36
PARA. (paragraph)	108.63	ENTITLE	44.3
ARTICLE	97.39	AGREEMENT	44.08
COURT	88.65	EVIDENCE	43.41
APPEAL	80.3	SENTENCE	42.39
APPELLANT	78.47	LIABILITY	41.65
LAW	73.55	ARGUMENT	41.29
JUDGMENT	71.67	GROUND	41.14
CLAIM	69.8	SUBMISSION	41.13
RIGHT	67.98	REGULATION	38.74
ORDER	64.39	CLAIMANT	38.32
DECISION	63.53	CRIMINAL	38.1
PERSON	62.83	REASONABLE	38.05
PROCEEDING	61.7	CONSIDERATION	38.02
DEFENDANT	57.72	NOTICE	37.82
PROVISION	57.55	PARTY	36.3
PRINCIPLE	55.77	DAMAGE	36.02
JURISDICTION	55.5	JUDICIAL	36
PARAGRAPH	54.69	MATTER	35.61
OPINION	54.4	EXTRADITION	35.54
APPLICANT	53.01	LORDSHIP	35.41
OBLIGATION	50.48	SUBJECT	35.02
ISSUE	50.22	UNLAWFUL	34.97
AUTHORITY	49.83	TERMS	34.26
REASON	49.66	ASYLUM	33.96
OFFENCE	48.8	STATUTE	33.72
FACT	48.36	PROPERTY	33.18
STATUTORY	47.47	REASONING	32.98
JUDGE	47.23	EFFECT	32.87
BREACH	47.16	LIABLE	32.46
CONCLUSION	46.37	CONDUCT	32.28
RESPONDENT	46.03	DECIDE	32.2
ACT	45.82	RELY	32.12
CONVENTION	45.71	DISMISS	31.75
RULE	45.67	TRIBUNAL	31.51
DUTY	45.15	CONTRACT	31.34
		ACCORDANCE	30.98

REVIEW	30.9	SUBMIT	23.47
ACCORDINGLY	30.18	HARM	23.31
LEGISLATION	30.11	SUFFICIENT	23.26
WITNESS	29.86	PRESUMPTION	23.17
LEASE	29.76	PREMISE	22.86
PROSECUTION	29.74	DISCLOSURE	22.79
IMPOSE	29.63	PROPOSITION	22.79
CLAUSE	29.53	PRISONER	22.76
SCOPE	29.08	JURISPRUDENCE	22.72
STATEMENT	29.01	CONSTITUTE	22.7
EXERCISE	28.83	DECEASED	22.69
DISCRETION	28.56	DETERMINE	22.69
REGARD	28.46	CONTRACTUAL	22.36
PARTICULAR	28.32	AUDITOR	22.16
TENANCY	28.32	SUBSTANTIVE	22.14
TRIAL	28.18	REINSURANCE	21.98
PROTECTION	27.98	DRAFT	21.95
COUNSEL	27.24	ESTOPPEL	21.9
DETENTION	27.15	ALLEGATION	21.76
PURSUANT	27.01	PUBLIC	21.61
PERMISSION	26.89	CONSPIRACY	21.54
CONVICTION	26.87	HEARING	21.44
IMPRISONMENT	26.82	BENEFIT	21.41
ACCOUNT	26.59	SUMMARY	21.32
INTEND	26.57	PLAINLY	21.29
FAIR	26.37	INDICTMENT	21.17
DOMESTIC	26.33	CONNECTION	21.09
JUSTIFY	26.08	PERSECUTION	21.02
CONCLUDE	25.96	FRAUD	20.99
TORT	25.7	ASSUMPTION	20.91
PROCEDURAL	25.32	INCOMPATIBLE	20.81
LEAVE	25.16	ARBITRATION	20.79
PLAINTIFF	25.1	DISCRIMINATION	20.7
COMPLY	24.95	CREDITOR	20.69
EXPRESSLY	24.93	LEGITIMATE	20.68
LEGAL	24.93	DETERMINATION	20.62
CONFISCATION	24.76	EMPLOYER	20.55
LAWFUL	24.31	PROPORTIONALITY	20.34
PURSUER	24.3	ALLEGED	20.3
INTERPRETATION	24.26	ENACT	20.19
INTENTION	24.22	VALUATION	20.09
ASSET	24.06	ACTION	19.95
AMEND	23.78	DISCLOSE	19.95
CONSTRUE	23.78	INCONSISTENT	19.95
NEGLIGENCE	23.77	CONSEQUENCE	19.85
COMMIT	23.66	SOLICITOR	19.67

PERMIT	19.56	TRAFFICKING	18.42
APPLICABLE	19.47	WORDING	18.38
LIMITATION	19.31	FORFEITURE	18.37
AMENDMENT	19.17	INTERFERENCE	18.36
COMPENSATION	19.17	DISPROPORTIONATE	18.28
FINDING	19.16	DISCHARGE	18.27
WARRANT	19.1	DEFENCE	18.22
ALLEGE	19.09	AMBIT	18.11
CONSTRUCTION	19.07	OBSERVE	18
FACTUAL	18.96	PRECLUDE	17.98
PREJUDICE	18.94	PROSECUTE	17.91
EXCEPTION	18.8	INVESTIGATION	17.81
PROPORTIONATE	18.77	ENACTMENT	17.72
DECLARATION	18.76	PAYABLE	17.72
LEGISLATIVE	18.62	INFRINGE	17.54

Chung (2003)

TERMS	TECHNICALITY		
		SCR	154.5577082
APPELLANTS	2012.581191	DICTA	151.8929201
PARA. (paragraph)	1444.695819	DEMISED	151.8929201
ESTOPPEL	975.3124342	FORFEITURE	148.6951744
LESSEE	639.5491372	JURISPRUDENCE	144.4315135
LAWFULNESS	563.6026772	TORTFEASOR	143.8985559
TORTIOUS	559.605495	REBUTTABLE	135.9041917
DISAPPLICATION	495.6505813	CONCURRING	135.9041917
FORESEEABILITY	439.6900318	APPELLATE	135.371234
INTERVENERS	439.6900318	DISAPPLY	127.9098274
SUBSECTION	384.7953975	ACCRUALS	127.9098274
NUPTIAL	373.736527	ILLIQUID	123.9126453
CHARTERER	351.7520255	HIRER	119.9154632
FREEHOLDER	327.7689328	LADING	119.9154632
LVT	319.7745686	EXTRADITION	118.5830692
RTS	317.7759775	TRESPASSERS	118.3165904
RESPONDENT	285.2534504	CONFISCATION	118.0985623
WAYLEAVE	255.8196549	IMPERMISSIBLE	116.7177175
OVERPAYMENTS	247.8252907	LORDSHIPS	113.2263013
ASSIGNEE	239.8309264	HDC	111.921099
CASSATION	239.8309264	COMPARATORS	111.921099
QUASHING	231.8365622	CONTRAVENTIONS	111.921099
ENFRANCHISEMENT	223.842198	UNREPORTED	109.5227897
DISCRIMINATOR	194.5295292	ADDUCE	103.9267348
SUBSECTIONS	189.8661501	NULLIFICATION	103.9267348
CONTROLEE	187.8675591	PROROGATION	103.9267348
BAILEE	159.8872843	UNAMENDED	103.9267348
GRANTOR	159.8872843	DEROGATING	103.9267348

AC	100.2268638	FORESEEABLY	55.9605495
COMITY	95.93237058	DPA	54.62815547
RESCISSION	95.93237058	DEMOTED	53.67644544
QC	94.08001789	UNJUSTIFIABLY	53.29576143
INTERLOCUTOR	93.26758251	CONTRACTUALLY	51.9633674
PROPORTIONALITY	93.26758251	SUBPARAGRAPH	50.63097336
LLP	92.73462489	COUNTERBALANCING	50.63097336
TORTS	89.53687921	AMBIT	50.63097336
UNPERSUADED	87.93800636	INCOMPATIBILITY	50.49072136
UNLAWFULNESS	83.94082426	PROPORTIONATE	50.46442411
WRONGDOER	83.94082426	EXERCISABLE	49.29857933
EXTRADITED	79.94364215	TENANT	49.2153047
PLAINT	79.94364215	CONSTRUE	48.96548082
SUBSTRATA	79.94364215	DEROGATE	47.96618529
CPO	78.34476931	DISALLOWANCE	47.96618529
CORONERS	77.27885408	DONEE	47.96618529
SEQ	77.27885408	REDELIVER	47.96618529
LIQUIDATOR	76.74589646	TRIABLE	47.96618529
OVERSUBSCRIPTION	75.94646004	VITIATE	47.96618529
RESIDUARY	74.61406601	SUBLETTING	47.96618529
JURISDICTION	74.55613583	CBL	47.96618529
ADMISSIBILITY	72.61547495	DISCLOSABLE	47.96618529
REINSURANCE	71.94927793	ENV	47.96618529
INTERLOCUTORS	71.94927793	FDS	47.96618529
JL	71.94927793	KB	47.56646708
ADJOURNAL	71.94927793	PURSUANT	46.87604471
DETERMINATIVE	70.35040509	INVESTIGATORY	46.82413326
DRAFTSMAN	69.28448986	PRECAUTIONARY	46.36731245
COMPULSORILY	68.75153225	ACTIONABLE	45.96759424
INCRIMINATION	67.95209583	TRESPASS	45.74552856
TORT	67.32096181	LIQUIDATED	45.68208123
LIQUIDATORS	66.35322298	CLAIMANT	45.59162166
AUDITOR'S	65.95350477	BREACH	45.58564128
ADJUDICATOR	65.18481591	JUDGMENT	44.57762762
CONSIGNEE	63.95491372	TENANCY	44.28623986
INJUNCTIVE	63.95491372	CONSCRIPTS	43.96900318
COMPARATOR	63.95491372	UNARGUABLE	43.96900318
INVALIDLY	63.95491372	FLAGRANT	43.33786916
CONTRAVENE	59.95773161	DETAINEE	42.96970766
ENFORCEABLE	58.18120623	INFRINGED	42.79336139
UNENFORCEABLE	57.10260153	PROSPECTIVELY	42.63660915
CLR	57.10260153	EVICTING	42.63660915
INFRINGE	55.9605495	RESETTLEMENT	42.37013034
ADOPTER	55.9605495	DEROGATION	42.25592514
INDICIA	55.9605495	PRECLUDED	42.1924778
CFR	55.9605495	CLAIMANT'S	41.97041213

PROCEEDINGS	41.67583522	SUBLET	37.307033
EXPRESSLY	41.57069392	CONTENDED	37.236907
JUSTICIABLE	41.57069392	QUASHED	37.026529
LBC	41.57069392	AVER	36.77407539
TENANCIES	40.44207779	OVERRIDDEN	36.77407539
FREEHOLD	40.29159564	DISHONESTLY	36.54566498
ARGUABLE	40.26790864	ARBITRATORS	36.54566498
TRESPASSER	39.97182107	UNLAWFUL	36.38460636
CONDUCE	39.97182107	PAROLE	36.37435718
DEBENTURE	39.97182107	CONTRACTUAL	36.09220315
IMPUTATION	39.97182107	BLAMEWORTHINESS	35.97463897
INCRIMINATE	39.97182107	COGENCY	35.97463897
APPEALABLE	39.97182107	RESCIND	35.97463897
ARBITRATIONS	39.97182107	UNEXPIRED	35.97463897
IMPRACTICALITY	39.97182107	COMMISSIONER'S	35.97463897
INAPPLICABILITY	39.97182107	ABSCONDED	35.97463897
MINUTER	39.97182107	APPEAL	35.42782883
SLR	39.97182107	LAWFUL	34.9997165
COURT'S	39.52768973	PRESUMPTION	34.88449839
PURSUERS	39.03130764	ADDUCED	34.64224493
VALUER	38.82976904	REBUTTED	34.64224493
FRAUDULENTLY	38.37294823	SEEKER	34.19811359
VALIDLY	38.37294823	PURSUANCE	33.97604791
JUDICIALLY	37.97323002	RESPONDENTS	33.90310663
REMITTED	37.83999062	SEVERALLY	33.5763297
INADMISSIBLE	37.307033		

Kit and Liu (2008)

TERMS	TECHNICALITY	LITIGATION	329.8128809
DETENTION	385.4495089	COMMISSIONER	329.5440896
UNLAWFUL	382.8202671	DISCLOSURE	329.2800942
ASYLUM	380.6266228	WARRANT	328.4183423
REASONING	372.86781	ASCERTAIN	327.029422
IMMIGRATION	365.5309131	SPOUSE	326.1300417
COUNSEL	364.9386094	ENTITLEMENT	323.2547026
SUBSTANTIVE	361.8189225	RESPONDENT	322.9890792
ACCORD	361.1647951	RETENTION	322.602551
CONTEND	356.4814787	CONFER	322.4806532
DETAIN	354.5130092	NATIONALITY	322.2783159
CONTRACTUAL	350.8853267	LEGISLATURE	320.0554374
INQUEST	350.1278397	CONSPIRACY	319.2519275
VIOLATION	349.5789043	DENIAL	317.4650382
WORDING	338.709086	PROSECUTE	314.9901571
RELIANCE	336.5960685	DONATION	311.6659089
PROCESSION	335.3939937	PROHIBIT	311.2845024

SUBMISSION	309.9496462	COMPLIANCE	253.549648
CONVICT	305.9522631	REFUGE	253.2367708
TERRITORIAL	303.3817378	TERMINATE	252.8597244
BENEFICIARY	302.0824348	DIRECTIVE	252.1124808
NEGLIGENCE	302.0766731	COMPULSORY	251.7042465
AMEND	301.6666066	COMPLY	250.3859527
DONOR	298.751768	SUE	250.3246864
PREJUDICE	297.7210468	JUDICIAL	250.2621231
UPHOLD	297.2745052	OFFENDER	250.1347118
MANDATORY	294.5363902	SANCTION	249.8839453
CLEARANCE	294.5186356	ACCUSED	249.4050624
IMPRISONMENT	294.294653	DULY	249.2161584
INJUNCTION	290.893216	DIVORCE	248.2623001
DISCHARGE	286.7826164	AUDITOR	247.2703
DISCIPLINARY	285.7922003	HARM	247.1610351
JURISDICTION	285.0844782	EXCEPTIONAL	247.0356603
ADVERSE	284.8597171	LIABLE	246.8289628
FRAUD	284.0492534	ACCORDINGLY	246.6060289
DEPRIVATION	281.9164668	CREDIBILITY	246.0638431
APPLICANT	279.0600525	PRECEDENT	245.9821236
ADVOCATE	276.9150856	CITIZENSHIP	245.9053654
CUSTODY	274.1102515	TERRORIST	245.5117178
UNREASONABLE	273.851935	PROSECUTION	244.9344922
LEGITIMATE	273.8410297	LEGISLATIVE	243.6737835
TRIBUNAL	271.6565193	BAIL	243.4332436
PROTOCOL	270.2046658	ALLEGE	243.1380383
APPLICABLE	269.9964232	REPAYMENT	243.0404791
SUICIDE	268.0610346	PRELIMINARY	241.9676258
PRESCRIBE	265.9456935	INAPPROPRIATE	241.7848922
EXPEL	265.83283	TARIFF	241.1493419
DISCRIMINATION	264.8033619	ALLEGATION	240.5880218
STATUTE	264.6616121	DISABILITY	240.5121299
PATENT	263.8735553	MANIFEST	240.3784216
IRRELEVANT	263.2935535	ATTEMPTED	239.9590167
CONTRARY	262.1948991	DEEM	239.8916796
VALUATION	261.658144	INCAPABLE	239.6615266
INTERIM	260.9920238	REMEDY	239.6551908
OFFEND	260.4276794	LEASE	238.9900365
DISCLOSE	259.5002352	BREACH	238.8770903
ALLEGED	258.4692269	DRAFT	238.8307584
ACCORDANCE	257.9307832	PRINCIPALLY	238.791215
COPYRIGHT	257.408554	ENJOYMENT	238.6273285
SCRUTINY	257.1653156	JUDGMENT	236.1902195
LEGALLY	256.5572938	REFUGEE	236.1144492
ENFORCEMENT	254.9869367	LIBERTY	235.4635556
FORMULATION	254.2681488	EXCLUSION	234.8602858

INTENT	234.857692	FAVOURABLE	221.1261115
HARDSHIP	234.8517479	ADMISSION	221.049458
KINGDOM	234.5582262	REFUSAL	220.7907886
DECREE	233.7414266	PRIVACY	220.7597222
WITNESS	233.5889369	COMPENSATION	220.7312108
JOINTLY	233.3680834	AMENDMENT	220.5299002
PAYABLE	233.1442494	CREDITOR	220.3557653
CONVICTION	232.8168463	VALIDITY	220.1608345
LANDOWNER	232.764052	PROCEEDING	219.8258206
UNFAIR	232.3539466	WARRANTY	219.7763427
JUSTIFICATION	231.8471873	COMMENCE	219.5055369
CONSTABLE	231.7783709	PROSPECTIVE	219.5014419
DETER	230.76392	PREVENTION	218.9361096
POSTPONE	230.1576089	RECORDER	218.8729502
STATUTORY	230.0926561	CONVENTION	218.6912044
INTERFERE	229.7940177	DECLARATION	218.513167
DESCENT	229.3528021	ENFORCE	218.4622437
RECONCILE	228.8839364	ASSURANCE	218.0703455
COMPETENT	228.6851807	SOLELY	217.5348873
INABILITY	228.4403583	PRESUME	217.2748821
INFLICT	227.9229806	SURRENDER	217.048841
ASSERTION	227.6907911	CONSISTENCY	217.0242153
DISPOSE	226.7592958	UNNECESSARY	216.9433013
REASONABLY	226.5074588	LANDLORD	216.8913344
POSSESSION	226.0632846	QUANTUM	216.6659721
CITE	225.8756556	SUBMIT	216.456664
RESTRAIN	225.47756	HEARING	215.9999566
RESIDENCE	225.4670075	THEREAFTER	215.8843558
SUSPEND	225.1567849	DISMISSAL	215.6949993
INTERCOURSE	224.6788679	SUMMARY	215.3576398
BINDING	224.4139673	CRIMINAL	215.3138359
OBLIGATION	224.375042	PARENTAL	214.8715328
LIMITATION	224.0288356	RESTRICTED	214.7517762
ENVISAGE	223.057449	JURY	214.6400222
ELIGIBLE	222.7895848	VERDICT	214.591902
COVENANT	222.7109117	APPRECIATION	214.4755396
DEFENDER	222.2255663		
PLEAD	221.4939192		

TF-IDF (Sparck-Jones, 1972)

TERMS	TECHNICALITY		
LAND	2636.403584	INVESTIGATION	1130.105264
ARTICLE	2538.995003	RELEASE	1124.777733
CONTRACT	2435.643508	CONSPIRACY	1118.490074
EXTRADITION	2278.499406	CLAIM	1109.252478
POSSESSION	2264.923951	APPLICANT	1105.006357
TENANT	2116.226356	SENTENCE	1099.965592
CONVENTION	2039.098672	DISABILITY	1092.526649
ASYLUM	1904.11018	TENANCY	1071.044383
DIRECTIVE	1847.999067	JFS	1069.334028
IMMIGRATION	1724.909458	TRAFFICKING	1059.828592
DISCRIMINATION	1702.053248	OFFENCES	1049.586658
SUICIDE	1696.919551	DUTY	1037.884671
ACCOMMODATION	1650.674541	EMPLOYMENT	1026.892183
CRIMINAL	1614.156299	TERRITORY	1003.458636
COMMISSIONERS	1600.183512	TAX	999.8874759
CLAUSE	1533.725826	PRISONER	968.8258019
PROPERTY	1525.313281	NEGLIGENCE	958.7472197
LEASE	1516.03417	CHIEF	952.2273471
CONFISCATION	1492.548537	PROCESSION	937.1691477
EHRR	1474.615217	CONVICTION	937.1604629
OFFENCE	1463.936264	DISCLOSURE	934.3572015
DAMAGES	1455.225399	SHAREHOLDERS	934.0637446
REGULATION	1449.019578	PROSECUTION	926.3526423
JURISDICTION	1437.269322	WITNESSES	925.5856864
REINSURANCE	1423.042814	DAMAGE	923.6596351
ASSETS	1408.93554	ESTOPPEL	923.5649161
DEFENDANT	1401.552443	BATTERY	922.7554111
SECRETARY	1390.300277	RIGHTS	921.6921067
ARBITRATION	1379.380848	DONATION	917.072036
SIAC	1359.796467	POWER	913.1745003
REGULATIONS	1354.555053	TORT	891.2326769
TRIAL	1246.984265	GOVERNMENT	876.8530247
LANDLORD	1246.459465	CREDITORS	873.633809
FRAUD	1244.940645	COUNCIL	870.3223428
DEFENCE	1241.474625	AGREEMENTS	860.5965536
REFUGEE	1230.626928	TREATMENT	860.4605404
TRIBUNAL	1227.156947	CROWN	859.1749194
WITNESS	1215.86235	SURVEILLANCE	855.5362617
LIABILITY	1205.104543	COMPENSATION	853.3622127
TENANTS	1181.945353	INQUEST	849.6045519
PERSECUTION	1167.605196	INJURY	846.4757987
PAROLE	1165.854401	POLICY	846.3394959
DETENTION	1155.073298	CRIME	840.5981644
AUDITORS	1153.679982	JURY	840.0957087
		UNLAWFUL	839.9769318

VALUATION	839.4769105	ASSURANCES	665.5349616
REFUGEES	836.7837236	MURDER	654.4989971
PATENT	826.6050837	BILL	650.6788265
BREACH	804.0211416	OBLIGATION	650.6051782
LICENCE	802.9947306	ADMISSIONS	638.2908403
SCHEME	802.8623434	ESTATE	637.9620713
CLAIMS	802.1752532	DIVORCE	636.4859145
DEFENDANTS	799.5934554	CONSTABLE	635.1519741
SUSPENSION	797.5254623	TRADE	635.1507899
ACCUSED	794.2504008	DISABLED	634.3984418
CERTIFICATE	793.2918534	AUTHORITY	634.1198409
SENTENCES	781.0619562	PROPRIETARY	633.1841368
TRUST	780.5829465	CHARTER	632.1923867
EVIDENCE	776.7241343	LEGISLATION	628.0044595
IMPRISONMENT	776.2555505	OFFICER	627.8952876
COMMISSION	775.0068924	DIVISIONAL	626.8234778
MEASURES	774.7352258	CONVICTED	626.4683268
ACTION	774.3413477	CONTRACTUAL	625.625533
PARTY	771.6210639	BUSINESS	624.6933944
FORFEITURE	768.7427279	CONTRACTING	622.6395109
PURCHASE	765.6485858	ARMED	620.7260616
CIVIL	764.6869899	LAWFUL	613.8232217
PERMISSION	760.8860037	JUDGE	613.6853061
COMPULSORY	748.8661396	APPLICANTS	611.9106552
CHARGES	748.8123612	PARENTAL	609.8689901
COMMISSIONER	747.8976241	VICTIM	608.7000314
IMGS	744.0607897	PROCEDURE	607.7223716
ADOPTION	743.7774835	AGENCY	605.940809
RULE	742.811319	STATUTORY	603.1450989
REGISTRATION	740.3564337	TERRORIST	598.5067965
ORDERS	735.3334967	PLAINTIFF	597.1491118
AGREEMENT	734.3044659	TRANSITIONAL	595.5205742
LIABLE	734.1028286	APPEALS	593.9366323
RULES	733.8337933	TERRITORIAL	593.3761971
INDICTMENT	727.3706272	REPORT	591.3430226
ADMISSION	726.1979316	WARRANT	590.9817738
DECEASED	720.1123125	CORPORATE	585.8757533
PURSUER	713.1783334	COMMITTED	585.6177201
CLAIMANT	711.3052988	PROOF	583.2636146
ADVOCATE	697.4993119	RESPONSIBILITY	582.9297225
COSTS	696.4864384	PRIVILEGE	580.7924329
CONDUCT	679.1330365	CHAMBER	578.9848803
FAIR	676.7580233	REGISTER	574.985382
SHERIFF	673.3189962	REMUNERATION	573.0717187
SENTENCING	671.789466	STATEMENTS	571.4209268
PRIVATE	669.0129893	JUDICIAL	569.349658

DEBT	567.8090598	LIBERTY	558.6016719
NOTIFICATION	567.577232	CORONER	557.2390365
ARREST	566.6545192	PURCHASER	556.8242929
TRUSTEE	566.2244824	JUSTICE	554.6968152
CUSTODY	565.2517359	AUDITOR	554.2493637
INSOLVENCY	565.0567297	DEFENDER	550.2867432
LAWS	564.265498	LOCALITY	548.13501
PROCEEDINGS	560.3327021	IPP	547.1739341
REQUIREMENTS	560.2252397	LIQUIDATORS	547.1739341

RIDF (Church and Gale, 1995)

TERMS	TECHNICALITY	CONTROLEE	4.382487027
ABA	0.992531351	DEMOTED	4.382487027
IMGS	6.263895489	EQS	4.377986469
AUDITOR	5.925572959	REMITTAL	4.377986469
HSMP	5.909177501	UKIP	4.377986469
ECRC	5.858308491	DISCRIMINATOR	4.340610458
MATRILINEAL	5.603109333	FFAS	4.314540446
LPP	5.557762462	CHARTERER	4.298102291
FOB	5.207024148	CONFISCATION	4.296192965
FSMA	5.207024148	HEARSAY	4.240999717
IPP	5.07592298	LLPS	4.17749656
LIQUIDATORS	5.07592298	REDELIVERY	4.17749656
DESCENT	4.973182839	LVT	4.175007168
WAYLEAVE	4.882050487	SPONSOR	4.163251131
REINSURANCE	4.861533021	CBL	4.103171899
ARBITRAL	4.858308491	DIC	4.103171899
ILLIQUID	4.839882726	EJECTION	4.103171899
SIAC	4.830875726	BARNARDISATION	4.024390816
BARNARDISED	4.796216162	GAK	4.024390816
PFT	4.796216162	LRR	4.024390816
BATTERY	4.793297209	ARBITRATION	3.970835022
INTERROGATION	4.655149866	DONOR	3.957568504
OVERDRAFT	4.60435413	INSURED	3.953193164
AQO	4.59605409	ACCRUALS	3.940612277
COPYRIGHT	4.59605409	RANSOM	3.940612277
DONATION	4.519213925	EIA	3.922811902
CONSIGNEE	4.496190681	INQUEST	3.894749691
CONSIGNOR	4.496190681	DEPORTEE	3.882050487
ADOPTERS	4.438451859	PAROLE	3.868807969
DDA	4.438451859	EJECTMENT	3.851190403
DEFAMATORY	4.438451859	HJ	3.851190403
SEISIN	4.435442445	COGNISANCE	3.839882726
UNDERTAKER	4.433609323	WAGES	3.833923519
FIXTURE	4.40927628	IRREDUCIBLE	3.822910281

DESTITUTE	3.796216162	ARBITRABILITY	3.284713952
AUDIT	3.785043837	NOTARY	3.284713952
CFR	3.755345485	PLAINT	3.284713952
DESTITUTION	3.755345485	PLENIPOTENTIARIES	3.284713952
JOINDER	3.741967217	ARREARS	3.275121259
SUICIDE	3.732866667	DIVORCE	3.269843679
DISAPPLICATION	3.728666198	ADOPTIVE	3.254920225
CHARTERPARTY	3.713139791	HRA	3.252522679
LESSEE	3.711309394	ACCUSERS	3.247822517
SUBLET	3.703968463	GRANTOR	3.247822517
UNDERLEASES	3.703968463	LACHES	3.247822517
DETERMINATE	3.681750081	DEBTOR	3.217490392
BAILMENT	3.655149866	LADING	3.211253662
SECONDMENT	3.652124264	FREEHOLDER	3.207024148
DEFENDER	3.62959632	INTERCOURSE	3.207024148
ICTA	3.551425873	JURE	3.188519992
PERSECUTION	3.549537956	OVERSUBSCRIPTION	3.17749656
ORDINANCE	3.54247228	PROPRIETARY	3.168349216
PENSIONABLE	3.540344257	MISFEASANCE	3.164077143
RESETTLEMENT	3.534360511	EXTRADITION	3.163622605
SURVEILLANCE	3.528880613	TRUSTEE	3.156719015
FLAGRANT	3.511258069	ADJOURNAL	3.136417749
PATENT	3.500898026	ANNUITY	3.136417749
EAW	3.496190681	SHERIFF	3.135571998
REQUISITION	3.486005238	IPPC	3.103171899
PURSUER	3.482334666	JL	3.103171899
LIQUIDATION	3.456540567	LIQUIDITY	3.103171899
ARBITRATORS	3.445472645	VINDICATORY	3.103171899
CREDITORS	3.444330268	TRADER	3.089732086
FORFEITURE	3.439947569	ADMISSIONS	3.087300214
ACCESSION	3.433609323	INSOLVENCY	3.059536408
TRAFFICKING	3.399787368	SUSPENSION	3.043354713
PREROGATIVE	3.388220338	ABIDE	3.03311965
RECTIFICATION	3.382487027	EXTRATERRITORIAL	3.024390816
CONSCRIPTS	3.377986469	PLEAS	3.013398467
DEMOTION	3.377986469	INSOLVENT	3.011091589
FORESIGHT	3.377986469	FIDUCIARY	3.005001623
TRANSNATIONAL	3.377288641	AIRSPACE	2.970202866
WARRANTY	3.377288641	DENOMINATION	2.970202866
TRESPASSER	3.36651438	HEREDITAMENT	2.970202866
CORONER	3.36195362	INVALIDLY	2.970202866
CARTEL	3.337849401	PEACEABLE	2.970202866
RESOLUTIONS	3.329147095	DEPUTE	2.966463373
REFUGEES	3.3180206	INHERIT	2.966463373
SEISED	3.302905084	PERSECUTED	2.960045818
HMRC	3.298102291	LEASE	2.952771118

TRESPASSERS	2.94170087	VICARIOUS	2.839882726
REGULATOR	2.940612277	PRECAUTIONARY	2.839272792
SLANDER	2.940612277	DIRECTIVE	2.838397813
TESTIMONY	2.925572959	LIQUIDATOR	2.82431378
FRAUDS	2.911228181	NOTIFICATION	2.822384702
REPURCHASE	2.911228181	DISABILITY	2.820624551
MATRIMONIAL	2.910574719	LIBEL	2.820163348
REFUGE	2.909177501	IMPAIRMENT	2.804738551
SENTENCING	2.903970066	FRAUDULENT	2.798256549
BCLC	2.901391693	COVERAGE	2.795754411
ACQUITTAL	2.891982635	INVESTIGATOR	2.793023968
CONSPIRACY	2.886154272	CONNIVANCE	2.781271134
VALUATION	2.885765568	RDC	2.781271134
CUSTOMARILY	2.885096053	SCRIVENER	2.781271134
TERRITORIAL	2.863871698	SETTLOR	2.781271134
HIRER	2.851190403	SURCHARGE	2.781271134
ILLEGALITY	2.850479945	VENTURERS	2.781271134
APPARATUS	2.839882726	MANSLAUGHTER	2.76740074
MALICE	2.839882726	REPAYMENT	2.764864701
OVERPAYMENTS	2.839882726	APARTMENT	2.755345485

Keywords (Scott, 2008)

TERMS	KEYNESS		
COURT	28955.793	PROVISIONS	4533.56982
PARA. (paragraph)	25311.1152	CASES	4428.68115
V. (versus)	22486.0918	BREACH	4419.3208
APPEAL	21236.8652	JUDGE	4404.61816
ARTICLE	19301.6328	APPELLANTS	4372.99072
ACT	18577.8652	PRINCIPLE	4212.11963
CASE	18328.9512	CRIMINAL	4197.90332
LAW	10458.0918	EHRR	4068.38989
JUDGMENT	9297.75	WLR	3907.81592
APPELLANT	8048.33496	POSSESSION	3887.15381
PROCEEDINGS	7787.61963	DUTY	3790.97339
CONVENTION	7764.64355	AGREEMENT	3788.57275
LJ	7707.0918	DEFENDANT	3774.44824
RIGHTS	7023.53613	QC	3683.96655
DECISION	6950.50488	APPLICANT	3642.31787
ORDER	6632.18164	AUTHORITY	3546.46826
JURISDICTION	6374.33105	SECRETARY	3524.19922
RELEVANT	6263.90625	FACTS	3516.82251
CLAIM	5832.43506	ISSUE	3374.36133
APPLICATION	5029.07129	EVIDENCE	3299.42969
CIRCUMSTANCES	4704.37988	RESPONDENT	3252.15674
STATUTORY	4629.16748	TRIBUNAL	3209.26611
		RULE	3033.38306

ENTITLED	2941.0791	CONDUCT	1515.09949
REASONS	2803.30688	DETENTION	1497.26697
EXTRADITION	2781.37329	DISCRIMINATION	1445.95166
TENANT	2751.19141	DISCRETION	1441.95447
JUDICIAL	2677.0271	DEFENCE	1389.27344
COURTS	2676.51587	CONSIDER	1386.31177
PARTIES	2654.61938	LEASE	1383.93884
OBLIGATION	2539.6167	INSTRUCTED	1371.33826
CONTRACT	2492.60645	DEFENDANTS	1366.88098
PROVISION	2477.90698	TERMS	1345.547
EWCA	2451.50879	PURSUANT	1339.63013
ARGUMENT	2386.76733	PROSECUTION	1328.84607
OFFENCE	2385.9021	CONFISCATION	1269.07068
ASYLUM	2349.10596	JUSTICE	1225.14673
UNLAWFUL	2339.28613	PROTECTION	1224.75037
DAMAGES	2309.95361	REQUIREMENT	1222.39429
LIABILITY	2284.2793	CLAIMS	1207.62708
REASONABLE	2278.59497	LAND	1197.73315
REGULATION	2119.13867	TENANCY	1195.66736
SENTENCE	2086.39966	IMPRISONMENT	1193.02539
REASONING	2025.00366	LAWFUL	1172.91772
COMMISSIONERS	2022.44434	TORT	1172.5343
IMMIGRATION	2010.72339	LANDLORD	1164.97839
CIV	2004.422	INTERPRETATION	1160.50037
GROUNDS	1989.08447	RULES	1157.02466
UKHL	1985.79382	ASSETS	1148.30078
CLAIMANT	1983.95984	COUNCIL	1144.61633
CLAUSE	1983.81494	SUBMISSION	1143.68054
QB	1928.78369	IMPOSED	1140.95483
OBLIGATIONS	1923.27112	DISMISS	1136.87476
PROPERTY	1895.76782	PROCEDURAL	1134.5509
RESPONDENTS	1884.95349	SIAC	1133.46716
LIABLE	1792.93555	LIMITATION	1131.70166
STATUTE	1745.29785	PERMISSION	1126.49219
DIRECTIVE	1661.50928	POWERS	1125.8125
REGULATIONS	1599.44641	PRINCIPLES	1120.33533
AUTHORITIES	1595.89124	ECHR	1120.28711
OFFENCES	1595.08813	REFUGEE	1093.42493
TRIAL	1585.17712	EWHC	1080.74683
LEGISLATION	1584.40833	JURISPRUDENCE	1076.47913
NOTICE	1575.83447	CONVICTION	1067.24646
ACCORDANCE	1555.87085	LEGAL	1063.48291
ALLEGED	1539.43848	ESTOPPEL	1049.19897
ACCORDINGLY	1533.99048	DRAFT	1037.41541
REASONABLY	1529.0011	ACTS	1036.97046
SUBJECT	1521.71167	DECISIONS	1024.77258

CONTRACTUAL	1009.71362	PROPORTIONALITY	800.885803
NEGLIGENCE	1008.14288	CONTRACTING	791.848633
FRAUD	1006.22614	SUBMITS	787.65509
APPELLATE	1003.68182	COMMISSIONER	787.418884
REINSURANCE	998.208862	HARM	786.077454
HEARING	996.033875	JFS	781.999695
WITNESSES	979.688477	WITNESS	778.102417
DISCLOSURE	977.463257	SUBSTANTIVE	777.979797
PRESUMPTION	940.262146	LEGITIMATE	776.405273
COUNSEL	917.103088	PROCEDURE	775.552551
REVIEW	914.585449	TENANTS	773.978577
ARBITRATION	913.435181	FAIR	771.727661
EMPLOYER	905.397217	ADVOCATE	765.725525
AMENDED	902.486267	PROPOSITION	763.004028
PERSECUTION	901.960632	FACT	756.85675
PAROLE	899.379883	INCONSISTENT	754.004761
DOMESTIC	899.247559	FORFEITURE	740.51178
SCHEME	895.321716	CONNECTION	739.816162
BAILII	878.652771	WARRANT	737.173157
ORDERS	877.052856	JUDGMENTS	734.561218
CONVICTED	873.508179	VALUATION	725.501587
INTENTION	873.116699	CROWN	717.422546
REMEDY	872.550049	DETERMINATION	708.487671
COMPENSATION	872.145813	PROPORTIONATE	708.239014
TRAFFICKING	870.371216	INDICTMENT	697.954224
PREJUDICE	861.558838	GROUND	697.223267
CPR	857.751099	PLAINTIFF	683.239563
REQUIREMENTS	843.486206	CONSIDERATIONS	679.011597
CONSPIRACY	838.205322	ASSESSMENT	674.228088
INCOMPATIBLE	836.546021	STATEMENT	670.859131
INVESTIGATION	806.776733		
DISMISSED	805.609436		

Terminus (Nazar and Cabré, 2012)

TERMS	WEIGHT		
EVIDENCE	4.341E+11	PURSUER	4.6433E+10
CLAUSE	2.4098E+11	OBLIGATION	4.5921E+10
CIRCUMSTANCES	1.7062E+11	ASSUMPTION	4.5656E+10
LAWFUL	1.1443E+11	PROVISION	4.4894E+10
SENTENCE	1.0438E+11	AUTHORITIES	4.2571E+10
ARGUMENT	8.4395E+10	PAYABLE	3.8687E+10
WITNESSES	7.1605E+10	BEHAVIOUR	3.6523E+10
DISCRETION	5.5265E+10	SUBMISSION	3.5257E+10
LAWFULNESS	5.0475E+10	INJUNCTION	3.5137E+10
PRESUMPTION	4.784E+10	ACCORDANCE	3.0595E+10
		PENALTY	2.9433E+10

JUDGMENTS	2.8225E+10	INSTANCE	1.0809E+10
COVENANT	2.591E+10	TRESPASSER	1.0741E+10
RETENTION	2.5783E+10	CONSIDERATIONS	1.0715E+10
CAUSATION	2.5584E+10	CREDITOR	1.0642E+10
SUBMISSIONS	2.3591E+10	LEGITIMATE	1.0298E+10
JOINDER	2.3264E+10	ALLEGATION	1.0296E+10
INFRINGEMENT	2.1797E+10	CONJUNCTION	1.027E+10
ENACTMENT	2.1287E+10	RELEVANCE	1.0087E+10
REQUIREMENT	2.0635E+10	TERMS	9915847757
PROHIBITION	2.0356E+10	ENTITY	9774607335
COMPLAINT	1.819E+10	UNDUE	9703178395
INTERFERENCE	1.7525E+10	TRESPASSERS	9615022449
DETENTION	1.7299E+10	MISFEASANCE	9387753592
PROVISIONS	1.7004E+10	ESSENCE	9357656006
OBLIGATIONS	1.6516E+10	ABIDE	9227638167
PROPOSITION	1.6505E+10	ARGUABLE	9221349763
JURISPRUDENCE	1.6399E+10	ENFRANCHISEMENT	9215991013
INDEMNITY	1.628E+10	INDICTMENT	9212504218
SEISIN	1.603E+10	NOTARY	9003557439
IMPUTATION	1.595E+10	DISCHARGE	8975964354
TENANCIES	1.5757E+10	DICTUM	8686269439
IMPRISONMENT	1.5671E+10	EXPIRY	8666756523
REMIT	1.5633E+10	OBJECTION	8450056594
INABILITY	1.5315E+10	MISCONDUCT	8448704896
AMBIT	1.5231E+10	VIOLATION	8410166552
ADJUDICATION	1.4645E+10	TARIFF	8317159790
ASSUMPTIONS	1.4467E+10	REGIME	8105979065
DISMISSAL	1.4269E+10	ENJOYMENT	7972251384
CONVICTION	1.4221E+10	DISMISS	7949027230
APPLICANTS	1.3976E+10	CERTAINTY	7822247098
REMITTAL	1.364E+10	REMISSION	7804842540
PRIVACY	1.332E+10	REVOCATION	7770253419
WARRANTY	1.2785E+10	NUPTIAL AGREEMENT	7763215309
OBITER	1.2671E+10	TENURE	7721706011
AVERMENTS	1.2555E+10	STATUTORY DUTY	7713722371
OWNERSHIP	1.2543E+10	FALSE	7546009023
SPOUSE	1.2195E+10	LEGALITY	7517004984
INADMISSIBLE	1.1933E+10	JURISDICTIONS	7377101293
DEPORTATION	1.1824E+10	SIGNIFICANCE	7194233887
DICTA	1.1483E+10	COMMITTAL	7129285212
COMITY	1.1427E+10	MOTION	7053415232
REASONING	1.1361E+10	AGENT	7034182349
LIABLE	1.1356E+10	ARGUMENTS	6939720151
JUSTIFICATION	1.1295E+10	APPLICABLE	6927901115
COGNISANCE	1.1112E+10	DISCRIMINATOR	6810030343
PREROGATIVE	1.0997E+10	RECOURSE	6709077432

CAPACITY	6567079153	EXCUSE	
CONTENTION	6562275925	CERTIFICATION	5373155311
PURSUANCE	6551920798	ADVERSE	5370927817
PRECLUDE	6495109059	DISHONESTY	5354468222
CONTEMPLATION	6400138987	HEREDITAMENT	5345809469
QUALIFY	6394824770	EX TURPI	5299999538
ENTITLEMENT	6379043323	IMMUNITY	5266701725
DETAINEE	6358644578	DEFENDER	5225341933
SUBSTITUTION	6332843693	LOCUS	5219165281
INQUESTS	6274645401	PROROGATION	5194875260
FIXTURE	6261603339	COLLUSION	5163692912
IRREDUCIBLE	6190152249	ASCERTAINING	5125620716
ASSIGNEE	6173211191	ATTRIBUTION	5086677440
NOTIFICATION	6169807658	DISPROPORTIONATE	5082196095
VICARIOUS	6112641179	INCOMPATIBILITY	5081480007
CULPABILITY	6081884638	EXPENSE	5038857630
CAUSAL	6077275277	STATUTORY PROVISIONS	5037514781
COMPLAINANT	6002287112	EXPULSION	4971130859
CONSIGNEE	5941667645	OWE	4964423345
PROCEED	5929419118	UNQUALIFIED	4859006745
PROPRIETOR	5910093720	LESSEE	4856801586
ASCERTAINMENT	5906252900	COMPETENCE	4845301187
PREMISE	5877544718	DOMESTIC LAW	4828508794
DURESS	5866461477	DILIGENCE	4772346867
ARREARS	5839363623	DISCUSSION	4758732771
INJURIES	5819606809	ASSIGNOR	4751146619
ALLEGATIONS	5815692882	REGARD	4743131546
DEPRIVATION	5751592459	PLEAD	4703802631
DENIAL	5734623945	CONFORMITY	4694710298
DWELLING	5714502607	SECURE TENANCY	4650784632
EXCLUDE	5707283864	TRANSACTION	4643735332
ADVICE	5704304397	REMITTED	4637282493
ADJUSTMENT	5686167071	ADMISSIBLE	4626556365
COHABITATION	5685346428	AMEND	4605381586
RESTRICTION	5649909218	POSSESSION ORDER	4564425281
FACTUAL	5639936135	CRIMINAL ACT	4557454015
AMENDMENTS	5555371714	CUSTODIAL	4557027720
PURSUIT	5534707984	TAXPAYER	4546474290
FAULT	5531743524	DECIDE	4538349121
CIRCUMSTANCE	5507258719	PURSUE	4530188439
CRIMINAL OFFENCE	5500545739	DEVOLUTION	4459793603
SUITABLE	5485371470	VICARIOUS LIABILITY	4407814998
IRRELEVANT	5455179985		
PUBLIC AUTHORITY	5393002876		
REASONABLE	5374556313		

229

C-value (Frantzi et al.)

TERMS	WEIGHT		TERMS	WEIGHT
COURT	1704.63145		HUMAN RIGHTS ACT	303.601888
CASE	1643.8973		GROUND	296.050836
SECTION	1372.89706		COMMON GROUND	276.695641
ACT	1112.81705		POWER	276.219736
HUMAN RIGHTS	1059.15024		INTERNATIONAL LAW	271.745091
APPEAL	1010.80826		EUROPEAN CONVENTION	270.273305
PARA. (PARAGRAPH)	945.881433		ENGLISH LAW	268.327143
ARTICLE	867.405675		SUPREME COURT	261.582763
LAW	856.308974		HIGH COURT	260.498961
COMMON LAW	770.945365		DEFENDANT	260.32976
STATE	753.63952		DIVISIONAL COURT	255.662491
ORDER	691.520055		OFFENCE	246.791788
DECISION	625.322923		CHIEF CONSTABLE	245.451346
MEMBER STATE	587.429663		JURISDICTION	241.091092
QUESTION	565.266781		RESERVED MATTER	240.837599
LOCAL AUTHORITY	553.853019		DRUG TRAFFICKING	237.119404
FACT	510.880296		ARGUMENT	235.23517
CLAIM	507.299529		REGULATION	234.723804
ISSUE	496.619524		CONCLUSION	233.946477
APPELLANT	478.427645		DUTY	233.410096
REASON	456.806618		PARAGRAPH	226.999042
RIGHT	434.325136		CONTRACT	226.774671
JUDICIAL REVIEW	409.792608		CRIMINAL LAW	226.050368
JUDGMENT	409.140172		GRAND CHAMBER	224.424962
EUROPEAN COURT	397.114441		SECURITY COUNCIL	219.276899
STRASBOURG COURT	396.044051		CONFISCATION ORDER	217.556631
RULE	394.969617		ACTION	216.538214
PROVISION	392.143775		APPLICANT	215.408238
PARTY	386.451689		PAROLE BOARD	211.937087
PRINCIPLE	371.071026		OBLIGATION	208.55757
APPLICATION	359.173002		DE FACTO	208.408767
PUBLIC AUTHORITY	357.991711		RESPONDENT	206.406841
DOMESTIC LAW	356.653724		SENTENCE	197.705192
TERM	355.873722		FAIR TRIAL	196.523481
CONVENTION	338.959318		ATTORNEY GENERAL	194.54326
OPINION	332.831099		PROPERTY	193.802944
JUDGE	328.802384		BREACH	188.667626
EVIDENCE	324.779823		POSSESSION	187.053491
MATTER	323.984365		REQUIREMENT	185.515989
CIRCUMSTANCE	322.635423		RELEVANT	184.646834
PROCEEDING	319.232693		DAMAGE	183.003678
AGREEMENT	317.919645		TENANT	181.09476
AUTHORITY	317.575501			

STATEMENT	178.813413	DEPUTY JUDGE	118.813215
POSSESSION ORDER	176.225007	GENERAL RULE	118.50739
APPELLATE COMMITTEE	173.046275	INPUT TAX	118.430933
BENEFIT	169.870871	PRIVY COUNCIL	117.564428
NUPTIAL AGREEMENT	166.980735	ENTRY CLEARANCE	117.408329
FIRST INSTANCE	164.639258	SPECIAL ADVOCATE	115.423649
STATUTORY PROVISION	163.259858	NATIONAL LAW	114.990396
NOTICE	158.81657	PROPRIETARY ESTOPPEL	113.363961
COUNTY COURT	156.582667	CONDUCT	112.550384
TRIAL	153.755174	CLAUSE	112.053188
DEFENCE	151.650073	SECURITY COUNCIL RESOLUTION	108.953903
TRIBUNAL	150.97505	CHIEF PLEA	107.038933
IMMIGRATION RULE	150.710817	LIFE SENTENCE	106.133219
UNLAWFUL MEAN	149.616642	ASYLUM	105.52197
SCHEME	147.870067	NATIONAL COURT	104.684076
GOVERNMENT	145.617113	REVIEW	103.957247
CRIMINAL PROCEEDING	145.037754	RELEVANT CHARGE	103.827765
AUTHORITIES	144.927971	CASE LAW	103.526718
FAIR COMMENT	144.859356	EUROPEAN ARREST WARRANT	103.29926
LIMITATION PERIOD	142.879711	DISTRICT JUDGE	102.510363
ENGLISH COURT	142.361781	LEAVE	102.409569
CARE AND ATTENTION	141.680881	POSSESSION PROCEEDING	102.173527
LEGISLATION	140.020277	LEGITIMATE AIM	101.901064
CONTROL ORDER	139.591361	STATUTE	99.6821522
HEALTH AND SAFETY	139.069252	CRIME	98.3522868
ARBITRATION AGREEMENT	138.9977	ACCOMMODATION	97.9723877
STATUTORY SCHEME	135.511289	ANONYMITY ORDER	96.9772731
TRIAL JUDGE	134.869055	EXTRADITION	96.6453338
JUSTICE	134.524555	CRIMINAL	95.8484439
CRIMINAL OFFENCE	132.300121	CHARGE	95.6606922
ASYLUM SEEKER	130.684806	LOSS OR DAMAGE	95.2156459
PROCEDURE	128.39408	DIRECTIVE	94.840907
COMMISSIONER	128.262215	EMPLOYER	94.8267722
INTER ALIA	127.37633	EXCEPTIONAL CIRCUMSTANCE	94.6529991
WITNESS	123.566645	FRESH CLAIM	94.1942609
RIGHTS AND FREEDOM	123.23625	DISCRIMINATION	93.1667693
HEARING DATE	123.094773	HEARING	92.1509329
CLAIMANT	122.003658	LANDLORD	91.9127213
LIABILITY	121.750034	ASSET	91.8399785
MATERIAL	120.444918	APPEAL COURT	91.4112486
STRASBOURG JURISPRUDENCE	119.494372	PROSECUTION	90.4156791
		CONSTRUCTION	90.3054392
		REASONABLE	90.2251337
		FUNDAMENTAL	89.9127035

RIGHT	
POINTE GOURDE PRINCIPLE	89.7747519
STATUTORY	89.1563388
INTERPRETATION	88.9339601
CRIMINAL JUSTICE	88.5953013
INVESTIGATION	88.5522693
REFUGEE CONVENTION	88.3071196
RELEVANT PROVISION	87.7719249
NOTIFICATION REQUIREMENT	87.0073611
POINTE GOURDE	85.784059
IMMIGRATION	85.4384484
REASONING	85.4039128
LAW COMMISSION	84.3823587
CONVICTION	84.0945537
ANONYMOUS WITNESS	82.9551729
REFUGEE	82.4712461

COMPULSORY ACQUISITION	82.182114
PRIMA FACIE	81.6312703
ANTE-NUPTIAL AGREEMENT	81.3495889
MATERIAL CONSIDERATION	80.2791996
PUBLIC AUTHORITIES	79.7440049
TERMS AND CONDITION	78.3488743
SPECIAL CIRCUMSTANCE	78.1384209
LEGITIMATE EXPECTATION	77.8708236
LEGAL ADVICE	77.1869638
COUNSEL	77.1534602
JUDGMENT GIVEN	77.0680316
FACTS AND ISSUE	76.7166061
ARREST WARRANT	76.6787991
SUBJECT MATTER	76.6398759

Termextractor (**Sclano and Velardi, 2007**)

TERMS	WEIGHT
HEARING DATE	0.8766511
APPELLANT	0.8734046
EUROPEAN CONVENTION	0.82952935
RESPONDENT	0.8214324
COMMON GROUND	0.7956776
ALLEGED	0.7838367
STATUTORY PROVISION	0.78352034
APPLICANT	0.78305364
PROCEDURAL	0.77347255
REASONING	0.7733148
WRITTEN SUBMISSION	0.7696053
DEFENDANT	0.769381
DOMESTIC LAW	0.76807594
LEGISLATION	0.76640904
ADMINISTRATIVE	0.7623122
APPELLATE	0.758357
COUNSEL	0.75561565
SOLICITOR	0.75396097
PARAS. (PARAGRAPHS)	0.74662244
TRIAL JUDGE	0.74570477
DECLARATION OF INCOMPATIBILITY	0.7452041
JURISPRUDENCE	0.7436813

MUTATIS MUTANDIS	0.74209946
PREMISE	0.74111587
CLAIMANT	0.7396065
BREACH OF ARTICLE	0.7392017
JUDICIAL REVIEW	0.7383706
ENTITLEMENT	0.73815125
AMENDMENT	0.733228
COURT OF APPEAL	0.7318615
LEADING JUDGEMENT	0.7260009
INQUIRY	0.7246992
CRIMINAL OFFENCE	0.7234405
VIOLATION OF ARTICLE	0.71955585
REASONED	0.71740055
QUESTION OF FACT	0.7169511
EXCEPTIONAL CIRCUMSTANCE	0.71629256
ORDINARY MEANING	0.7153073
FINDING OF FACT	0.71429193
JUDICIAL DECISION	0.71283436
CENTRAL ISSUE	0.71115977
REQUIREMENT OF ARTICLE	0.710623
GROUND OF APPEAL	0.71027756
STATUTORY POWER	0.7097792
PRIMA FACIE	0.7059722

RULING	0.7058605	STATEMENT OF FACT	0.6701049
PUBLIC AUTHORITY	0.7034754	AMBIT OF SECTION	0.66928416
MATTER OF PRINCIPLE	0.7022409	FACTUAL BASIS	0.6684911
CLAUSE	0.7020176	WORDING	0.6680128
ALLEGATION	0.7018192	ORAL SUBMISSION	0.66800904
NATIONAL COURT	0.7005917	EUROPEAN COURT	0.6675826
JURISDICTION	0.69970083	CRIMINAL TRIAL	0.6675294
DOMESTIC COURT	0.69937414	ADJUDICATOR	0.66713107
ORAL ARGUMENT	0.69914114	MANDATORY	0.66706246
CONSENT	0.6989916	INJUSTICE	0.6669288
STATUTORY SCHEME	0.6988523	ORAL EVIDENCE	0.6656444
WITNESS STATEMENT	0.69809383	WIDE MEANING	0.66525626
IMPRISONMENT	0.69573575	LIMITATION PERIOD	0.66436666
STATUTE	0.6929306	INTERIM	0.6634462
ADVISER	0.69230694	INTERVENER	0.6625426
LEGITIMATE AIM	0.6919335	SUPERVISION	0.662202
ASSERTION	0.6909509	COMPELLING REASON	0.66153353
TENANT	0.690775	ENFORCEABLE	0.6612625
LITIGATION	0.6902687	ISSUE OF PRINCIPLE	0.6610027
CONSEQUENT	0.6901087	STANDARD OF PROOF	0.6609384
STATUTORY DUTY	0.68993646	FACTUAL BACKGROUND	0.66023993
SUBSIDIARY	0.68993205	LANDLORD	0.6601711
CIVIL PROCEEDING	0.68946004	ENFORCEMENT	0.66011816
PLEA	0.68812364	LEGAL SYSTEM	0.658957
PROSECUTOR	0.6875431	MAGISTRATE	0.65863883
RELEVANCE	0.68743384	AMBIT OF ARTICLE	0.65770876
DISCLOSURE	0.6874094	UNFAIR	0.65752345
INSTANCE	0.68708616	PERSUASIVE	0.65597165
ACCUSED	0.6861996	COUNTY COURT	0.6555662
EXPRESS PROVISION	0.6859706	RISK OF HARM	0.65396476
DECISION-MAKING PROCESS	0.685206	SUITABILITY	0.65378207
INVESTIGATION	0.68454194	MEMBER STATE	0.65359306
REVENUE	0.68373346	ENGLISH LAW	0.65309894
INFERENCE	0.68370485	APPELLATE COMMITTEE	0.6529512
PROSECUTING	0.68352455	DECEASED	0.65161127
CONTRACTING	0.6829005	PETITION	0.65021
CERTAINTY	0.68264705	INCONSISTENCY	0.64739084
IDENTIFIABLE	0.68239766	PRACTITIONER	0.6469096
BURDEN OF PROOF	0.68144417	PERMIT	0.64683723
IMMIGRATION	0.6803734	DISSENTING OPINION	0.64628696
CRIMINAL PROCEEDING	0.67985916	CUSTODIAL SENTENCE	0.6460633
ASYLUM SEEKER	0.678483	CRIMINAL ACTIVITY	0.64590555
AMBIT	0.67680043	DEGRADING TREATMENT	0.6452102
ENACTED	0.6746165	TREATY	0.6447396
INTERNATIONAL LAW	0.6743448	ADMISSIBILITY	0.6446799
CRIMINAL CONDUCT	0.6730207	SURROUNDING CIRCUMSTANCE	0.64453304

CRIMINAL CHARGE	0.6440158	MARGIN OF APPRECIATION (MOA)	0.63030034
ORAL HEARING	0.6438487	SENTENCE OF IMPRISONMENT	0.6299473
EXTRADITION	0.6436387	PARAMOUNT CONSIDERATION	0.6292698
PRELIMINARY ISSUE	0.6435879	REASONABLE DOUBT	0.62912965
QBD	0.64357597	AVOIDANCE OF DOUBT	0.6279483
SPOUSE	0.6431005	REFUSAL OF LEAVE	0.6278775
CERTIFIED	0.64305156	SECURITY OF TENURE	0.6274126
FACT-FINDING	0.6422165	CONTRACTING PARTY	0.62701553
ISSUE OF FACT	0.642073	ADMISSIBILITY DECISION	0.6249697
INFRINGEMENT	0.6415607	UNLAWFUL ACT	0.62457865
DEPRIVATION OF LIBERTY	0.6409908	SKELETON ARGUMENT	0.62452304
INTENT	0.6409419	SEXUAL ASSAULT	0.6226731
ILLEGITIMATE	0.6408357	LEGAL POSITION	0.62260467
TRANSACTION	0.6404267	LEGAL ADVISER	0.62259203
TENANCY	0.6401971	REASONABLE GROUND	0.6217673
LEGAL PROCEEDING	0.639156	LOCAL HOUSING AUTHORITY	0.62032735
MEMBERSHIP	0.6378892	BREACH OF CONTRACT	0.6202623
CONCURRENT	0.63753384	PRIVY COUNCIL	0.6172742
CONTRACTUAL OBLIGATION	0.63674057	CLAIM FORM	0.6172292
CATEGORY OF PERSON	0.63673466	PARENTAL RESPONSIBILITY	0.61704516
CRIMINAL ACT	0.63639355	PRINCIPLED BASIS	0.6163145
DECLARATORY RELIEF	0.6358489	LEGAL PRINCIPLE	0.6160962
ALLEGEDLY	0.63565975	CRIMINAL LAW	0.6154744
ASYLUM CLAIM	0.63484645	LEGISLATIVE HISTORY	0.6147502
ADJUDICATION	0.6334796	PROCEDURAL FAIRNESS	0.6147446
JUSTICE SYSTEM	0.6328674	ULTRA VIRES	0.6146327
ENGLISH COURT	0.6323865	DRAFTING HISTORY	0.61326534
DOCUMENTATION	0.6323153	PRELIMINARY POINT	0.6119381
BALANCING EXERCISE	0.63227826	ESSENTIAL POINT	0.61136866
CHANGE OF CIRCUMSTANCE	0.63227355	DUTY OF CARE	0.6112761
CERTIFIED QUESTION	0.6310944		
SUBORDINATE LEGISLATION	0.63059807		

Textract (Park et al., 2002)

TERMS	WEIGHT		
RIGHTS	1100.88565	SEISED	5.49358837
CONTROLEE	6.33182093	PARA. (PARAGRAPH)	5.47476774
HLR	5.75157662	CHARTERPARTY	5.42970234
LLP	5.68523377	SCR	5.27416342
APPELLANT	5.56934907	DEPORTEE	5.25045524
DISCRIMINATOR	5.5277393	MISFEASANCE	5.06585248
IAT	5.51335255	WAYLEAVE	5.04962604
HDC	5.50669104	AVERMENT	5.04024155
		REDELIVERY	5.01845348

LAWFULNESS	4.99923095	OVERPAYMENT	3.81537536
GAK	4.96756534	COMPARATOR	3.80045084
NASS	4.9582435	TRESPASSER	3.78559793
CHARTERER	4.9298921	SUBSECTION	3.77129242
DEROGATING	4.84319792	JURE	3.73048635
CIRCUMSTANCE	4.82501304	DISAPPLY	3.72501922
BCLC	4.69097044	IRREBUTTABLE	3.72015533
TRAVAUX	4.67241417	REBUTTABLE	3.7157473
IMPERMISSIBLE	4.67034074	ALLEGATION	3.71326204
PROCEEDING	4.65835179	JUSTICIARY	3.70119514
CONSPIRATOR	4.64803274	HEREDITAMENT	3.67684661
RESPONDENT	4.62451452	INTENDMENT	3.67653486
JOINDER	4.57833416	DISJUNCTIVELY	3.65223106
UNCONSCIONABILITY	4.57019564	FORFEITURE	3.64643783
JURISPRUDENCE	4.56937614	PAROLE	3.63739447
CLAIMANT	4.56911896	COMPLAINER	3.63391657
REPRESENTEE	4.50437927	WRONGDOER	3.62134736
PURSUER	4.47524885	CONSIGNOR	3.56245953
DPA	4.42319183	ACTIONABILITY	3.55591733
ESTOPPEL	4.40652866	DETERMINATIVE	3.54757769
SEISIN	4.40431188	LEASEHOLDER	3.54334143
CASSATION	4.38181409	FRAUDSTER	3.53585043
DISAPPLICATION	4.37780543	JUDGMENT	3.52682324
DETAINEE	4.3719146	ADMISSIBILITY	3.51477961
OVERSUBSCRIPTION	4.37008721	JURISDICTION	3.51167644
CPO	4.33601528	EJECTMENT	3.50707561
EXTRADITION	4.28797425	COMITY	3.5013808
FORESEEABILITY	4.21870788	DECLARATORY	3.49766982
NULLIFICATION	4.20167956	FLAGRANT	3.48788474
CONFISCATION	4.192229	PATENTEE	3.4834512
SEEKER	4.17153502	FREEHOLDER	3.482147
UNLAWFULNESS	4.11560018	COUNTERBALANCING	3.46843105
TORTFEASOR	4.04040582	ACTIONABLE	3.46412331
PROROGATION	4.03759016	TRAFFICKING	3.45044454
PRESENTMENT	3.93544873	ASSIGNOR	3.4490584
DANGEROUSNESS	3.92737812	PROBATIVE	3.44022726
COGNIZABLE	3.92083592	PURSUANT	3.437451
CONTUMACY	3.92083592	DEROGATION	3.43375596
HOTCHPOT	3.92083592	JUSTICIABLE	3.41207052
PROPORTIONALITY	3.90953982	AMBIT	3.4088118
CIF	3.89753692	TORT	3.40534581
AVER	3.8968065	ASYLUM	3.40447916
APPLICANT	3.89518239	CONCURRING	3.40361764
DISENTITLE	3.85846467	ENFRANCHISEMENT	3.40079654
ADOPTER	3.82963791	EXECUTRIX	3.39591917
INTERLOCUTOR	3.81806488	UNCONSCIONABLE	3.39359686

PERPETRATOR	3.39244666	LAWFULLY	3.16056306
INCOMPATIBILITY	3.3801761	EXTRADITE	3.15712066
HUMAN RIGHTS	3.37182824	UNLAWFULLY	3.14688761
ADJUDICATOR	3.36069741	EIA	3.13894336
UNREPORTED	3.35453946	SERVICEMAN	3.12587085
ORDINANCE	3.3448941	ABET	3.12520221
COMPLAINANT	3.3428113	GRANTOR	3.12320055
LIQUIDATOR	3.33458804	RECIDIVIST	3.12170151
EVICTING	3.33290325	TENANCY	3.11091243
INDICTMENT	3.32696181	LAWFUL	3.10729258
QUASHING	3.32657548	SECTION	3.10508982
REFUGEE	3.31908858	IRREDUCIBLE	3.10182511
REQUISITION	3.31453334	PRESUMPTION	3.09889404
CONVENTION	3.31208506	TENANT	3.08819693
COMMISSIONER	3.30360732	INADMISSIBLE	3.08767972
COGNIZANCE	3.30119114	DETENTION	3.08321556
APPEAL	3.29771516	INSURER	3.07834626
ILLIQUID	3.29448547	APPEALABLE	3.06903766
TRIBUNAL	3.29143486	INDICIA	3.05650571
DDA	3.29008414	DELICT	3.05588515
ACCUSER	3.2894499	TERRITORIALITY	3.05342444
INAPPLICABILITY	3.28302406	HEARSAY	3.03728261
FOB	3.28088883	LEGATEE	3.03429679
ADMINISTRATRIX	3.27481859	VIOLATION	3.03356165
UNENFORCEABILITY	3.27481859	DISPROPORTIONATE	3.02947649
OBLIGATION	3.27124738	DECLARATOR	3.02839859
PERSECUTION	3.27076688	DILATORINESS	3.02839859
EXTRADITED	3.27034109	HAPPENSTANCE	3.02839859
CPA	3.25918761	REASONING	3.02608169
DEPORTATION	3.25153698	AUDITOR	3.00966627
DEBATEABLE	3.23714079	DEPORT	3.00794338
PARAGRAPH	3.23674006	COMPLICIT	3.00282543
REGULATION	3.23591445	EXTRATERRITORIALITY	3.00209657
EXCULPATORY	3.23463221	MOA	3.00209657
BLAMEWORTHINESS	3.232282	OFFENDER	2.9834043
IMMIGRATION	3.23096067	LEGALITY	2.96837543
COGNISANCE	3.22807752	LESSEE	2.95984569
GRANTEE	3.22453733	BLAMEWORTHY	2.95479657
CONSCRIPT	3.22401357	PARTICIPATOR	2.94604015
REALISABLE	3.21502466	LAUNDER	2.94176283
DEFENDANT	3.21054146	OFFENCE	2.93608739
COMPULSORILY	3.20968975	PERADVENTURE	2.93208486
BAILMENT	3.19121864	COURT	2.9293644
UNLAWFUL	3.18355769	CONSTRUE	2.91495577
ACQUITTAL	3.16633104		
ENACTMENT	3.16091744		

TermoStat (Drouin, 2003) MWTs

TERMS	WEIGHT
COMMON LAW	36.46
PUBLIC AUTHORITY	28.15
JUDICIAL REVIEW	27.25
DOMESTIC LAW	25.75
COMMON GROUND	21.53
INTERNATIONAL LAW	20.44
DUTY OF CARE	19.69
MARKET VALUE	19.4
CAUSE OF ACTION	19.19
STATUTORY PROVISION	18.7
MEMBER STATE	18.61
POSSESSION ORDER	18.27
DE FACTO	17.21
CRIMINAL LAW	16.67
ASYLUM SEEKER	15.71
FAIR COMMENT	15.7
CASE LAW	15.36
SPECIAL ADVOCATE	14.64
NATIONAL LAW	14.3
PRIMA FACIE	14.01
CRIMINAL OFFENCE	13.97
LEGITIMATE AIM	13.73
LEGITIMATE EXPECTATION	12.19
PUBLIC LAW	12.19
STATUTORY DUTY	12.19
EX PARTE	12.03
STATUTORY RIGHT	12.01
DECLARATION OF INCOMPATIBILITY	11.95
ARREST WARRANT	11.91
FUNDAMENTAL RIGHT	11.69
BREACH OF CONTRACT	11.66
PROCEDURAL OBLIGATION	11.26
DIRECT DISCRIMINATION	11.02
STATUTORY POWER	10.91
PRIVATE LAW	10.87
QUESTION OF FACT	10.84
QUESTION OF LAW	10.75
SERIOUS INJURY	10.53
PERSONAL INJURY	10.45
FINDING OF FACT	10.43
CRIMINAL CONDUCT	10.27
FAIR HEARING	10.21
INDIRECT DISCRIMINATION	10.16

REASONABLE DOUBT	9.81
CIVIL LAW	9.79
CRIMINAL ACT	9.56
TAXABLE PERSON	9.43
LAW OFFENCE	9.22
CONSTRUCTIVE TRUST	9.11
POSSESSION CLAIM	9.11
UNLAWFUL ACT	8.81
NOTICE OF DEFAULT	8.77
CRIMINAL STANDARD	8.77
CRIMINAL COURT	8.47
LEGAL RIGHT	8.1
REASONABLE EXCUSE	8.07
ARBITRATION CLAUSE	8.06
JOINT ENTERPRISE	8.04
UNLAWFUL DISCRIMINATION	7.79
CIVIL COURT	7.79
RACIAL DISCRIMINATION	7.78
LEGAL SYSTEM	7.66
DERIVATIVE CLAIM	7.55
VILLAGE GREEN	7.39
CUSTODIAL SENTENCE	7.34
BODILY HARM	7.25
PRO RATA	7.2
QUANTUM MERUIT	7.01
REASONABLE TIME	6.84
CIVIL RIGHT	6.62
GRAND JURY	6.31
NUPTIAL AGREEMENT	6.26
CIVIL LIABILITY	6.19
LEGAL PRIVILEGE	5.93
FALSE IMPRISONMENT	5.89
CRIMINAL CHARGE	5.16
ADOPTIVE PARENT	5.08
RESTRAINT OF TRADE	5.02
EXCLUSIONARY RULE	5.02
CIVIL ACTION	4.93
ROYAL PREROGATIVE	4.77
RIGHT OF WAY	4.65
JUDICIAL PROCEEDING	4.65
PUBLIC DOMAIN	4.62
CIVIL CASE	4.37
NATURAL JUSTICE	4.37
CONFLICT OF INTEREST	3.84

238

```
┌─────────────────────────────────────────────┐
│                                               │
│        PRIVY COUNCIL. UNITED KINGDOM          │
│                                               │
└─────────────────────────────────────────────┘
```

Privy Council Appeal No 95 of 2006.
Quincy Todd, *Appellant,*
v.
The Queen, *Respondent,*
FROM
THE COURT OF APPEAL OF
THE BAHAMAS.
- - - - - - - - - - - - - - - -
JUDGMENT OF THE LORDS OF THE JUDICIAL
COMMITTEE OF THE PRIVY COUNCIL,
Delivered the 8th April 2008.
- - - - - - - - - - - - - - - -
Present at the hearing:
Lord Hope of Craighead
Lord Rodger of Earlsferry
Lord Walker of Gestingthorpe
Lord Mance
Sir Christopher Rose.
- - - - - - - - - - - - - - - -
Delivered by Sir Christopher Rose:

1. The appellant appeals by special leave of the Board, given on 22nd November 2006. Following a 4 week trial in the Supreme Court of the Bahamas, before Dame Joan Sawyer CJ and a jury, he was convicted on 23rd January 1998, of the murder of the deceased, Venette Bellizaire, in 1994. He was sentenced to death. On 16th December 1999 the Court of Appeal of the Commonwealth of the Bahamas (Carey P, Zacca and Hall JJA) dismissed his appeal.

2. The deceased was killed some time after 6.30am on 25th August 1994, when she was seen getting into a uniquely distinctive Nissan motor car which belonged to Otis Palmer. Mr Palmer gave evidence that he lent the car to the appellant between 9pm and 10pm on 24th August and he returned it at about 8am on 25th. Mr Palmer's girlfriend Margaret Barr gave evidence that the car was missing when she returned home about 1am on the morning of the 25th and the appellant returned it later that morning. Later that day she found two earrings in the car which she handed to the police on 21st September. The deceased's parents identified the earrings as belonging to her. Mr Palmer, meanwhile, had been arrested on 20th September on suspicion of the murder, but he was released following the appellant's arrest. This took place at 5.20am on 21st September and, on the same day at 1.20pm, he was shown the earrings by D.S. McCoy. Later the same day, in circumstances giving rise to the first ground of appeal, the appellant went with police officers to bushes near Cromwell Drive, Freeport. Bones, said to be those of the deceased and showing fractures of the skull due to several blows, were found, together with her slippers and a bangle.

3. During the trial, while D.C. Johnson was giving evidence of taking photographs, on 21st and 22nd September, defence counsel indicated that, at the end of cross-examination, he would be making submissions about two of those photographs, numbered 4 (of the appellant pointing at the skeletal remains) and 25 (of the appellant pointing at a garbage bin). It was put to D.C. Johnson in cross-examination that, following the shooting of a snake by D.C. Wilchcombe, the appellant had been forced at gunpoint to kneel and point for photograph 4 and that the following day, the appellant had again been forced to point for photograph 25. He denied these allegations.

4. At the conclusion of D.C. Johnson's evidence, in the absence of the jury, defence counsel submitted that photographs 4 and 25 should be excluded as being more prejudicial than probative and because they had been obtained by improper means, contrary to s.178 of the Bahamian Evidence Act 1996. The judge pointed out that there was no evidence of improper means. Counsel said he would return to the matter on the voir dire. Other evidence was then called for the prosecution including that about the motor car, the last sighting of the deceased and the earrings to which reference has already been made.

5. A voir dire was then held to determine the admissibility of oral and written confessions said to have been made by the appellant. Several police officers gave evidence in relation to events at the police station following the appellant's arrest and at the crime scene. In particular, D.S. McCoy gave evidence that the appellant said at the police station "Otis Palmer who is in the cell have nothing to do with it. I killed her and I can show you where it happen". He directed the police towards a bushy area at Cromwell Drive and said "I killed her over there". When they reached the bushes the accused pointed and said "There are the bones and clothing of Venette Bellizaire". The appellant gave evidence denying making these statements and saying that a written confession (the terms of which were not seen by the trial judge) was induced by mistreatment by a number of officers: he was threatened, beaten, pistol whipped and given electric shocks. He had deliberately misspelt his name in the written statement to prove that he had been beaten. The photographs taken at the crime scene were staged. The photographer was already there on 21st when he arrived with other officers. After a snake had been shot at several times and killed by D.C. Wilchcombe, he was threatened with being shot and was forced to point for photograph 4 and, the nextday, for photograph 25. He called his brother and Otis Palmer in support of the allegations of mistreatment, all of which were denied in evidence by the several officers said to have been involved.

6. The judge ruled that the oral and written confessions be excluded. She said she was not sure (the onus being on the prosecution) that no threats were made. She had a reasonable doubt about whether the snake-shooting incident occurred. She was not satisfied the appellant had been told that he could consult an attorney, as is guaranteed by Article 19(2) of the Bahamas Constitution. She said "It doesn't take much to make a statement inadmissible". A discussion took place with Counsel about the status of the evidence about the visits to the crime scene on 21st and 22nd. Reference was made to sections 20 and 178 of the Evidence Act. A distinction was drawn between the confessions and the visits. The judge said "Obviously he's been to the Cromwell Drive area because that's apparent from the pictures and that's where the police say they found the bones. To some extent there's a conflict between them as to how that was done. Police say he pointed it out and he says he didn't. So obviously the jury will have to resolve that. I don't know that I can do very much about that because the pictures have gone in now you see". Later, the

judge said to prosecuting counsel "The point is, anything that suggests he made a confession is basically out. But of course the exception is – so much of the confession as relates to the finding of these things is admissible. That's what sub-paragraph 5 says" (She was clearly intending to refer to s.20(4) of the Evidence Act). She also referred to <u>Lam Chi-ming v The Queen</u> [1991] 2AC 212 and to the difference between s.20(4) and the English Police and Criminal Evidence Act 1984, (PACE) and the Indian Evidence Act 1872. She referred to s.178 but concluded that, in view of the terms of s.20(4), she could not exclude the evidence about the finding of the bones under s.178(1).

7. In due course, when summing up, the judge posed the question in relation to events at the crime scene "Who do you believe, Mr Johnson or the accused and the other police witnesses who said he led them to the spot?"

8. After the judge's ruling, the jury were recalled. D.S. McCoy gave evidence in chief about going to the crime scene at the accused's direction and the taking of the photographs. He denied, in cross-examination, knowing in advance where the bones were. He made no reference in his evidence to the confessions which the judge had excluded. Subsequently, at the jury's request, the court went to view the crime scene. Two weeks after he had first given evidence before the jury, D.S. McCoy was recalled, apparently for the purpose of putting on record what had taken place during the view. His evidence in chief, which again included an account of how the accused had shown where the bones were so as to explain what the jury were shown, passed without incident. In cross-examination he was challenged about what had happened when he first went to the scene with the accused and, in particular, about the direction in which the accused had pointed. His answer included the following: "when I stopped the car he pointed to the western side….and said 'Venette, I killed her over there'. We exit the car and he led us to the bushes." Defence Counsel said "I'm only speaking about where the accused pointed. He never told you anything." The judge intervened: "The jurors will disregard any answer about what the accused is supposed to have said. There's no evidence before you about that." Cross-examination continued about other matters.

9. The appellant's defence was alibi, in support of which he called his brother-in-law. He claimed to have been at a family prayer-meeting between 6am and 7am on 25th August and not to have borrowed Palmer's car on that date.

10. The Court of Appeal held that evidence of what the appellant did leading to discovery of the bones was admissible under the terms of s.20(4) although this did not relieve the judge of the obligation to decide admissibility by reference to s.178. The Court concluded that the trial judge had exercised her discretion under s.178 but, if they were wrong, they would have exercised the discretion so as to admit the evidence.

11. Section 20(1) and (2) of the 1996 Act, which are, with immaterial differences, identical to s.76(1) and (2) of PACE, provide for the admissibility of confessions proved beyond reasonable doubt not to have been obtained by oppression nor rendered unreliable by anything said or done at the time. Section 20(4) provides: "The fact that a confession is wholly or partly excluded in pursuance of this section shall not affect the admissibility in evidence of any facts discovered as a result of the confession <u>and of so much of the confession as relates thereto</u>." Save for the additional words underlined, this provision is identical to s.76(4)(a) of PACE. Section 76(5) and (6) of PACE are the entirely different provisions which the judge had in mind: They render a fact discovered as a result of an excluded confession

12. admissible only when evidence is given by the accused or on his behalf as to how the fact was discovered.

12. Section 178(1) (like s.78(1) of PACE) provides: "In any criminal proceedings the court may refuse to allow evidence on which the prosecution proposes to rely to be given if it appears to the court that, having regard to all the circumstances, including the circumstances, in which the evidence was obtained, the admission of the evidence would have such an adverse affect on the fairness of the proceedings that the court ought not to admit it."

13. Before the Board, Mr Guthrie QC (who did not appear below) advanced two grounds of appeal. First, the judge having excluded the oral and written confessions ought also to have excluded the evidence of the appellant directing the police to and at the crime scene. Secondly, the giving of inadmissible evidence by D.S. McCoy should have resulted in the discharge of the jury.

14. As to the first ground, he accepted that trial counsel should have challenged the admissibility of the evidence before he did. He submitted that s.20(4) does not authorise the admissibility of any part of an involuntary confession. The fact that the bones were discovered was admissible but any part of the inadmissible confessions and the evidence as to why the police went to the crime scene were inadmissible. He relied on the authority of Warickshall (1783) 1 Leach 263 and also on two authorities concerned with s.27 of the Indian Evidence Act 1872 Kottaya v Emperor AIR 1947 PC 67 and Anter Singh v Rajashthan 2004 ILR 1.543. The judge should have excluded evidence as to the appellant's conduct leading to the discovery of the bones as being part of the excluded confessions. Counsel referred to Ibrahim v The King [1914] AC 599 and Lam Chi-ming v The Queen (above). He submitted that if s.20(4) lays down a different rule from these authorities it must be mitigated by s.178, otherwise it will be inconsistent with the Constitution of the Bahamas, which prohibits torture and inhuman or degrading treatment and guarantees a fair hearing. He referred to Bowe v The Queen [2006] I WLR 1623, Pillay and others v S (2004) 2 BCLR 158 and Sweeney [2000] 50 OR (3d) 321. In saying what she "must" do by reference to s.20(4) the judge showed she wrongly believed she had no discretion.

15. As to the second ground, Mr Guthrie submitted that, following D.S. McCoy's gratuitous evidence, which had been ruled inadmissible, the judge should have discharged the jury. Counsel relied on her later comment that she would have done so had the evidence been given in chief.

16. Mr Dingemans QC (who did not appear below) on behalf of the Crown submitted, as to the first ground, that the photographs did not advance the argument: the important matter was the direction of the police to the scene by the accused. When summing up, the judge, in clearest terms, left for the jury's resolution the conflict between the police and the appellant as to why they went to the crime scene and whether the photographs were staged. He referred to Gould (1840) 9 C&P 364 as the genesis of s.20(4): a policeman to whom a statement was made under peculiar circumstances was permitted to state, after a lantern had been found, that the prisoner had told him that he had thrown it there. The English common law developed along different lines from the Bahamian legislation – see Liam Chi-ming v The Queen (above) and Timothy v The State [2000] I WLR 485. The plain meaning of s.20(4) is that parts of a confession shown to be true by a subsequent discovery are admissible. There has been such a legislative provision in the Bahamas since 1904. There are similar provisions in the legislation of other countries including India (s.27 of the Indian Evidence Act 1872 – see Anter Singh v

Rajasthan (above)), Ceylon (s.27 of the Evidence Ordinance – see Ramasamy [1965] AC1) and South Africa (s.218 of the Criminal Procedure Act 1977 – see Pillay & Others above) none of which has been the subject of effective constitutional challenge. Case law in Ontario has developed a similar principle (see St Lawrence [1949] OR 215 and Wray [1971] SCR 272) albeit with a modification providing a discretion to exclude, arising from the Canadian Charter of Rights and Freedoms (see Sweeney (above) paragraph 2(c) and (d) of the judgement).

17. Mr Dingemans further submitted that, the confessions having been excluded because the prosecution could not prove they were voluntary, the evidence about the appellant directing the police to the scene was properly admissible under the clear words of s.20(4). In so far as s.20(4) might reward police misconduct, whereas modern democracy requires fair play, s.178 operates to afford the judge an overriding discretion. Had the judge been sure that the snake-shooting incident had occurred she could only properly have excluded the evidence that the appellant took the police to the bones. As she was not sure, she was entitled to admit that evidence in order to give effect to s.20(4) in the light of s.178. It is a proper exercise of discretion to admit what the jury might well conclude has been proved to be true but to exclude what cannot be so proved.

18. As to the second ground of appeal, Mr Dingemans submitted that the judge dealt with D.S. McCoy's answer immediately and properly, and no submission was made to discharge the jury, even when the judge later made the comment about what she would have done had the evidence been given in chief. Continuing with the trial was an unassailable exercise of discretion.

19. In the light of these submissions their Lordships, in disagreement with the Court of Appeal, do not accept that the trial judge exercised a discretion in relation to the evidence which she admitted under s.20(4). The language of her ruling shows that she believed she had no such discretion. The Court of Appeal said they would have exercised discretion, if the judge had not, in favour of admitting the evidence. The question which now arises is whether had she realised she had a discretion, the judge would inevitably have exercised it to admit the evidence.

20. It is apparent that at trial objection to the admissibility of evidence about the visit to the crime scene was not made when or in the way in which it should have been. Defence Counsel should have raised the matter and sought a voir dire before evidence of the photographs was led. The focus of his complaint should have been not on the photographs but on the circumstances whereby the appellant and the police officers came to the crime scene. Photograph 25, of the appellant pointing at the garbage bin, proved nothing and was therefore irrelevant and inadmissible. Photograph 4 of the appellant pointing at the ground was clearly stage-managed to some extent; no good reason is apparent for taking it; it added nothing to the case; and it should not have been admitted. But the photographs in themselves, were of no significance and their admission in evidence cannot have had any adverse impact on the fairness of the trial or the safety of the jury's verdict. The crucial question, on this first ground of appeal, is whether the evidence of the accused directing the police to the crime scene and the bones was properly admitted in the light of sections 20(4) and 178(1).

21. The words of s.20(4) are plain. Their Lordships are of the view that they mean what they say, namely, facts discovered as the result of an excluded confession and so much of the confession as relates thereto are admissible. The words of the subsection were not challenged before the Board on the ground of unconstitutionality. Nor, so far as is known, have similar provisions in other

jurisdictions been subjected to successful constitutional challenge. But the 1996 Act must be read as a whole and, in particular, s.20(4) read in the light of the over-riding discretion which s.178(1) confers on the trial judge.

22. In their Lordships' judgment, evidence of the discovery of the bodily remains at the appellant's direction (if the police evidence was accepted), was admissible within s.20(4) notwithstanding the exclusion of the prior written and oral confessions. The judge, in her ruling, made no finding of police impropriety. She expressed doubt about whether the snake-shooting incident had occurred and whether the accused had been told of his constitutional right to representation. Those doubts were sufficient to render the oral and written confessions inadmissible under s.20(1) and (2), as she rightly ruled. Their Lordships are of the view that, had the judge realised that she had a discretion under s.178(1), she would, inevitably, in the absence of any proved impropriety have exercised it in favour of admitting the evidence and so have left it to the jury (as she did) to determine whether the conditions of s.20(4) were satisfied, in that, as the prosecution contended, the deceased's bones were discovered as a result of the appellant's confession. The first ground of appeal therefore fails.

23. As to the second ground, the alleged admission inappropriately rehearsed by D.S. McCoy was, clearly, potentially damaging to the defence. But the context in which the evidence came to be given and the events which followed it are important. There is no reason to believe that the witness deliberately gave evidence which he knew had been ruled inadmissible. No such allegation was made by counsel at the time. There is nothing to suggest that the trial judge thought this was the case. When D.S. McCoy first gave evidence before the jury about events at the crime scene, he made no mention of the admission. He referred to it only when cross-examined when giving evidence before the jury for the second time following the court's view at the scene. Furthermore, when defence counsel made his submission of no case to answer, he referred to D.S. McCoy having "slipped when he said certain things to the jury".

24. The critical question is whether, the inadmissible evidence having been given, the judge's failure to discharge the jury rendered the trial unfair or the verdict unsafe. Immediately the answer was given, the judge directed the jury to disregard it. Furthermore, that direction was plainly heeded because the judge, at a later stage, described having seen the jury nod in agreement with her direction. No one suggested at the time that the jury should be discharged. It is true that, during the submission of no case, the judge commented that she would have discharged the jury had the inadmissible evidence been given in chief. Whether she would have taken that view after hearing submissions on the matter, it is impossible to say. Even at that stage, having heard the judge's comment, no counsel suggested that the jury be discharged. Following the judge's direction to disregard it, no further reference to the inadmissible evidence was made during the trial. In the Court of Appeal trial counsel did not argue this ground.

25. In their Lordships' view, an appellate court, remote from the atmosphere and nuances of the trial process, should be slow to interfere when a trial judge continues with a trial after the jury has heard inadmissible evidence and will not do so merely because it would have decided differently. In this case the judge immediately, and effectively, directed the jury to disregard the evidence; no further reference was made to it; and none of those involved in the trial appears to have thought that the evidence was so damaging that consideration should be given to discharging the jury. The trial was in its third week and all the evidence was completed on the

following day. It is clear that the judge thought about what she was doing and must have been satisfied that the jury would be able to return a proper verdict. The local appeal court upheld her decision.

26. In these circumstances, it is impossible to conclude that the trial was unfair or the verdict unsafe. The second ground of appeal therefore fails.

27. Their Lordships will humbly advise Her Majesty that the appeal against conviction should be dismissed. The Crown concedes that the mandatory sentence of death cannot stand. Their Lordships will further advise Her Majesty that the appeal against sentence should be allowed, the sentence of death quashed and the case remitted to the Supreme Court of the Bahamas for determination of the appropriate sentence.

JUDGMENT

Allison (Appellant) *v* Her Majesty's Advocate (Respondent) (Scotland),
before
Lord Hope, Deputy President,
Lord Rodger
Lord Walker
Lord Brown
Lord Kerr.
JUDGMENT GIVEN ON
10 February 2010
Heard on 8 December 2009.

LORD RODGER:

1. On 9 September 2004 the appellant, Steven Allison, was convicted after trial in the High Court at Glasgow of four contraventions of section 4(3)(b) of the Misuse of Drugs Act 1971. In effect, he was found guilty of being concerned in the supplying of cocaine and three other controlled drugs at his home in Cumbernauld, at an address in Falkirk and elsewhere in the United Kingdom, between 12 November and 3 December 2003. The trial judge, Lord Bracadale, sentenced him to 8 years imprisonment.

2. The appellant appealed against both his conviction and sentence. On 7 November 2008 the appeal court (Lord Osborne, Lady Paton and Lord Philip) refused his appeal against conviction, leaving his appeal against sentence to be heard on a date to be fixed.

3. Among his grounds of appeal against conviction was one which was first advanced in an additional Note of Appeal. It relates to the record of a police interview of a John Stronach. Mr Stronach had died before the trial and the Crown introduced the interview into evidence in accordance with the procedure in section 259(5) of the Criminal Procedure (Scotland) Act 1995.

4. Neither before nor during the trial did the Crown disclose to the defence that Mr Stronach had a number of previous convictions and outstanding charges. In particular, he had convictions for reset, theft by opening lockfast places, assault and robbery and assault and breach of the peace. He also had a number of outstanding charges, including two alleged contraventions of the Misuse of Drugs Act 1971, an alleged theft by housebreaking and several alleged contraventions of the Road Traffic Act 1988. One of the outstanding cases under the Misuse of Drugs Act related to events covered by the trial and was known to the appellant's legal advisers. The Crown disclosed the previous convictions and the other outstanding charges only while the appellant's appeal was pending before the appeal court. This prompted the appellant to lodge his additional ground of appeal: "The failure on the part of the Crown to disclose to the defence the existence of all the previous convictions and outstanding charges resulted in the defence being unable to prepare and properly conduct their defence and the result was that the appellant did not receive a fair trial, as guaranteed by article 6(1) of the European Convention on Human Rights."

5. Following the dismissal of his appeal by the appeal court, the appellant applied for leave to appeal to the Privy Council in relation to the additional ground of appeal. On 6 March 2009 the appeal court (Lord Osborne, Lady Paton and Lord

Mackay of Drumadoon) refused the application as incompetent, on the ground that no intimation of a devolution issue had been given to the Advocate General as required by para 5 of Schedule 6 to the Scotland Act 1998. The court went on to indicate that, if it had been open to them to grant or refuse leave, they would not have granted leave.

6. The appellant subsequently applied to the Privy Council for special leave to appeal. The Board granted special leave. Although the statement of facts and issues included an issue relating to the competency of the appeal court's decision to refuse leave, neither the advocate depute nor the Advocate General advanced any argument on the point at the hearing of the appeal. Undoubtedly, when the appeal court determined that the Lord Advocate was not under an obligation by virtue of article 6(1) of the European Convention to disclose the outstanding charges against Mr Stronach, they were in substance determining a devolution issue in terms of para 1(d) or (e) of Part I of Schedule 6 to the Scotland Act 1998 – irrespective of whether all the relevant procedural steps had been followed. It follows, as was held in *McDonald v HM Advocate* 2008 SLT 993, 1002, paras 48 and 49, that an appeal on that point lies to this Court under para 13(a) of Part II of that Schedule.

7. Of course, the late Mr Stronach's name was never included in the list of Crown witnesses appended to the indictment for the appellant's trial – which may help to explain why the need to disclose his criminal antecedents was overlooked. But, when dismissing the appellant's appeal, the appeal court rightly accepted, under reference to *Holland v HM Advocate* 2005 1 SC (PC) 3, 24, para 72, that the failure by the Crown to disclose Mr Stronach's previous convictions had been incompatible with the appellant's article 6(1) Convention rights. Despite the further conclusion of the Privy Council in *Holland*, at pp 24-25, paras 73-74, that the Crown were also under an obligation to disclose material outstanding charges of which they were aware, the appeal court in the present case drew a distinction between Mr Stronach's previous convictions and "his outstanding cases at the time of the trial" – by which the court obviously meant the charges against him which had been outstanding at the time of his death. The court continued:

> "We consider that, in this context, a distinction has to be made between previous convictions and outstanding cases. While, in appropriate circumstances, the existence of previous convictions may be of importance in connection with the preparation of a defence and to the challenge that may be mounted to the credibility of a witness, we do not consider that the same may be said of outstanding cases. Where an individual is charged with crime, he or she is presumed to be innocent until proved guilty. If a case is outstanding, necessarily no verdict has been reached in it. In these circumstances we have insuperable difficulty in understanding how information relating to those matters could be properly deployed in the conduct of a defence."

8. Previous generations of Scots lawyers and judges do not appear to have experienced the same "insuperable difficulty" as the appeal court.

9. It is, of course, trite that an individual charged with crime is presumed to be innocent until proved guilty. But that is not to say that he has to be treated in all respects as if he were an innocent person against whom no charge had been brought. Most obviously, in an appropriate case, he can be remanded in custody

pending trial or granted bail subject to appropriate conditions. Similarly, depending on the offence and the terms of his contract of employment, he may be suspended from his employment. More generally, if you know that someone has been charged with, say, fraud, you will be less inclined to enter into a commercial transaction with him; if you know that someone has been charged with sexual abuse, you will think twice before entrusting your children to her care; if you know that someone has been charged with theft, you will be less inclined to trust anything which he tells you, unless it can be confirmed from other sources.

10. The Privy Council's decision in *Holland*, that the Crown should disclose outstanding charges of Crown witnesses of which they were aware, simply reflected the common sense position that - just as in everyday life - judges or jurors who have to assess the credibility of a witness may properly take into account not only the fact that the witness has been convicted of various offences, but also the fact that he has been charged with others. To judge from the passage quoted in para 7 above, the appeal court seem to have thought that this was an unprincipled and incoherent innovation. It is noteworthy that they did not refer to any authority. In reality, the approach of the Privy Council, in so far as it proceeds on the basis that outstanding charges may have a bearing on a witness's credibility, merely reflects what appears to have been recognised as the proper practice in Scottish courts for more than 170 years.

11. At one time, in Scots law anyone convicted of serious crimes became technically infamous (infamis) and was thereafter unable to give evidence at any trial. By the early nineteenth century this rule was proving self-defeating for the authorities: all too often it was a gift to the defence since it prevented the only material witnesses to crimes from giving evidence. So the rule was first relaxed and then eventually abolished. The only explicit authority relating to a witness with outstanding charges comes from that era. At a High Court trial at Dumfries, when leading a Crown witness, William Higgins, the advocate depute began by establishing that he was due to be tried at the same circuit on a charge of theft by housebreaking, aggravated by his having been previously convicted of theft and being a thief by habit and repute. See *John Hannah and Hugh Higgins*, 17 September 1836, Bell's Notes, p 256, in the Supplement to Hume's *Commentaries on the Law of Scotland respecting Crimes* (1844 edition), vol 2. Since the court ruled on the admissibility of the advocate depute's line of questioning, the defence must have objected that the Crown were, in effect, leading a witness who, if convicted of the crime in question at his trial later in the sitting, would then be unable to testify. The court rejected the argument and allowed the question. As the author of the Notes, Sheriff Bell, comments, "The court, however, in allowing the question, must have thought it relevant to affect the credit of the witness."

12. The potential relevance of outstanding charges to the credibility of a witness appears to have been settled in the nineteenth century. See, for instance, Dickson's *Treatise on the Law of Evidence in Scotland* (revised edition, 1887) vol 2, para 1619. Most significantly, Macdonald's *Criminal Law of Scotland* (3rd edition, 1894), p 462, says: "Nor may [a witness] refuse to say whether he has been convicted of or stands indicted for a crime." This passage appears in virtually the same words in the fifth and final edition (1948), pp 297-298. The passage could never have stood unchallenged in successive editions of the traditional vademecum of Scottish criminal practitioners and judges if it had not

reflected practice in the courts. Not surprisingly, therefore, neither the advocate depute nor the Advocate General supported the approach of the appeal court in the present case.

13. In *Holland v HM Advocate* 2005 1 SC (PC) 3 the appellant was convicted of a charge of assault and robbery at a house in Rutherglen. The Crown failed to disclose that there were outstanding charges against the complainers, relating to drug dealing at the house in question. The Privy Council held, at p 25, para 75, that information about these charges would have helped to complete the picture both of the complainers and of their milieu. In other words, it would have had the potential to weaken the Crown case and so it should have been disclosed. In his written submissions in the present appeal, the advocate depute gave examples of other situations where an outstanding charge against a Crown witness might materially weaken the Crown case or strengthen the defence case: if the witness denied he had ever been in trouble with the police, an outstanding charge could legitimately be put to him; similarly an outstanding charge might provide a potential motive for the witness giving untrue information in an attempt to curry favour with the authorities.

14. It is unnecessary to prolong the discussion since the Crown did not deny that the outstanding charges against Mr Stronach might have weakened the Crown case by casting doubt on his character or credibility. It follows that, in accordance with *Holland v HM Advocate* 2005 1 SC (PC) 3 and *HM Advocate v Murtagh* 2009 SLT 1060, the failure of the Crown to disclose the outstanding charges to the defence was indeed incompatible with the appellant's article 6(1) Convention rights.

15. At the hearing of the appeal, all this really went without saying and the only live issue was the actual significance, in the whole circumstances of the case, of the Crown's failure to disclose the charges. The appeal court did not consider that matter, but they did, of course, consider the effect of the Crown's failure to disclose his previous convictions. Having considered the circumstances, the appeal court were not "persuaded that the failure of the Crown to disclose the previous convictions of Stronach to the appellant's advisers resulted in an unfair trial and hence a miscarriage of justice." They accordingly rejected the appellant's appeal, so far as based on the Crown's failure to disclose Mr Stronach's previous convictions.

16. Standing that decision, at the hearing before this Court, Mr Jackson QC, who appeared for the appellant, had to argue that the failure to disclose Mr Stronach's outstanding charges made a significant difference. In other words, the Court should conclude that there would have been a real possibility of a different outcome if the jury had been made aware, not only of Mr Stronach's previous convictions, but of the outstanding charges against him: in that event, the jury might reasonably have come to a different view as to whether the appellant was concerned in the supplying of the various drugs during the relevant period.

17. The case against the appellant was circumstantial. It comprised, for the most part, evidence of observations by police officers who had conducted a surveillance operation over several weeks. The evidence relating to Mr Stronach's statement concerned events of 24 November 2003.

18. DS Duncan Smith was not otherwise involved in the relevant events. He gave evidence that, at about 12.45 pm on 24 November, when checking an address, he saw a blue Peugeot, registration number M810 UEW, parked at the appellant's home at 58 Whitelees Road, Cumbernauld. At 9.32 pm Mr Stronach was seen

driving the Peugeot to a service station at Kilmarnock where he met up with a Ford Orion. The two cars drove in convoy to Logan, near Cumnock. There, in the car park of the Logangate Arms, the driver of the Orion spoke to Mr Stronach who did not leave his car. Mr Stronach then drove up to Glasgow and on to the M8 where he was stopped by two police officers. A Farm Foods bag, found in the glove compartment of the Peugeot, contained cocaine worth at least £30,000.

19. The police interviewed Mr Stronach on tape in the early hours of 25 November. This is the interview which is the subject of the additional ground of appeal. The tape recording of the entire interview was played to the jury during the evidence of DC McFadden. In the course of the interview Mr Stronach said that he had been sent by a man called "Stevie" from Abronhill to sell the car. The appellant's first name is Steven and his home was in the Abronhill district of Cumbernauld. The description of "Stevie" given by Mr Stronach fitted the appellant. On 27 November the appellant left his home and drove to the house of Mr Stronach's girlfriend in Denny. He then took Mr Stronach's passport to Airdrie Sheriff Court where it was used in connexion with his application for bail. A receipt for the passport from the court dated 27 November was recovered from the appellant's home.

20. Defence counsel took DC McFadden through the transcript of the interview in detail and was able to show that Mr Stronach had told many lies. When he came to address the jury, the advocate depute accepted that he had clearly lied about his movements and about his involvement in drugs. But the advocate depute suggested to the jury that it would be easier to accept those parts of the interview which were supported by other acceptable evidence. In particular, he pointed to the evidence of DS Smith, who was not otherwise involved in the investigation, that the Peugeot which Mr Stronach was driving when stopped by the police had been parked outside the appellant's house earlier the same day. The advocate depute also referred to the evidence about the appellant collecting Mr Stronach's passport from his girlfriend's house and taking it to Airdrie Sheriff Court in connexion with his application for bail on the drugs charges arising out of the recovery of the cocaine from the Peugeot. The advocate depute argued that it would be a spectacular coincidence if this did not indicate that the appellant knew of Mr Stronach's involvement with drugs.

21. In his supplementary report to the appeal court, Lord Bracadale, summarised the position in this way:

> "Taking into account the analysis of the interview of Mr Stronach carried out by [defence counsel] and the concessions made as to his credibility by the advocate depute, the jury would have been most likely to conclude that Mr Stronach did indeed tell many lies in the course of the interview. They would, however, have been entitled to be selective in their view of the evidence of Mr Stronach."

Lord Bracadale then referred to Mr Stronach's previous convictions and added:

> "In the circumstances outlined above it is difficult to see how the canvassing of the previous convictions of Mr Stronach before the jury would have bolstered the already largely successful attack on his credibility. It is also difficult to see why knowledge of the previous convictions would have discouraged the jury from being

selective in the approach to the contents of the interview of Mr Stronach."

22. Against that background, the appeal court were not persuaded that the failure of the Crown to disclose Mr Stronach's previous convictions resulted in an unfair trial and hence a miscarriage of justice. For exactly the same reasons, I am not persuaded that, if defence counsel had been able to deploy Mr Stronach's outstanding charges as well as his previous convictions, this would have made any material difference. More especially, it would not have affected the fact that the jury, who must have been well aware of the defects in Mr Stronach's statements, could still, with equal plausibility, have accepted those elements, and only those elements, in Mr Stronach's account which were corroborated by other acceptable evidence.

23. I am accordingly satisfied that there is no real possibility that the jury would have come to a different verdict on the four charges against the appellant if they had been made aware, not only of Mr Stronach's previous convictions, but of the outstanding charges against him as well. There has therefore been no miscarriage of justice. I would accordingly dismiss Mr Allison's appeal and remit the case to the appeal court to proceed as accords.

LORD HOPE:

24. I agree with Lord Rodger that the appeal must be dismissed, and I would make the same order as he proposes.

25. The point of principle which this case raises is whether a failure to disclose outstanding charges against a Crown witness is incompatible with the accused's article 6(1) Convention rights. Had it not been for the passage in the opinion of the appeal court which Lord Rodger has quoted in para 7 of his judgment, I would not have thought that there was now any room for dispute on the point. In *McDonald v HM Advocate* [2008] UKPC 46. 2008 SLT 993, para 51 Lord Rodger said that the decisions of the Board in *Holland v HM Advocate* [2005] UKPC D 1, 2005 SC (PC) 3 and *Sinclair v HM Advocate* [2005] UKPC D 2, 2005 SC (PC) 28 had answered this question. Included within the general description of disclosable material are two classes of material, namely police statements of any witnesses on the Crown list and the previous convictions and outstanding charges relating to those witnesses.

26. The rule of law on which that classification is based is that of fairness. In *McLeod v HM Advocate (No 2)* 1998 JC 67, Lord Justice General Rodger said that our system of criminal procedure proceeds on the basis that the Crown have a duty at any time to disclose to the defence information which would tend to exculpate the accused. In *Sinclair v HM Advocate*, para 33 I said that the prosecution is under a duty to disclose to the defence all material evidence in its possession for or against the accused, and that for this purpose any evidence which would tend to undermine the prosecution case or to assist the case for the defence is to be taken as material.

27. Sometimes the proposition is worded differently. In *HM Advocate v McDonald* [2008] UKPC 46, 2008 SLT 993, para 50 Lord Rodger said:

"Put shortly, the Crown must disclose any statement of other material of which it is aware and which either materially weakens the Crown case or materially strengthens the defence case (disclosable material)"

Lord Bingham of Cornhill used the same formula when describing the "golden rule" in *R v H and others* [2004] UKHL 3, [2004] 2 AC 134, para 14 when he said:

> "Fairness ordinarily requires that any material held by the prosecution which weakens its case or strengthens that of the defendant, if not relied on as part of its formal case against the defendant, should be disclosed to the defence."

In *HM Advocate v Murtagh* [2009] UKPC 36, 2009 SLT 1060, para 11, I said, under reference to *McLeod*, *Holland*, *Sinclair* and *McDonald*, that it was well settled that the Crown must disclose any statements or other material of which it is aware which either materially weakens the Crown case or materially strengthens the case for the defence: see also Lord Rodger, para 48.

28. These formulations should however be regarded as expressing what has been described as the golden rule in shorthand. After all, they are describing a decision about disclosure which must normally be taken before the trial. It is a decision which will be based on an assumption as to what may happen in the future. So the question the Crown must ask itself is what the possible effect would be likely to be if the material were to be disclosed. As I said in *R v Brown (Winston)* [1998] AC 367, 374, it would be contrary to the principle of fairness for the prosecution to withhold from the defendant material which *might* undermine their case against him or which *might* assist his defence. Lord Collins, referring to what I said in that case, also used the word "might" in *Murtagh*, para 75. That is the way Lord Rodger has expressed the position that the Crown has adopted in this case in para 14, above, and I respectfully agree with it.

29. As for the point that troubled the appeal court, it is true that a distinction can be drawn between previous convictions and outstanding charges. But that does not mean that it can be assumed that information about outstanding charges of Crown witnesses can never affect their credibility. It is enough, for the disclosure rule to apply to them as a class, that they *might* do so. Of course the person concerned is presumed to be innocent until proved guilty. But if he is asked the question whether he has ever been in trouble with the police, he must answer it. A false or evasive answer might well be thought by a jury to undermine his credibility. Other circumstances may be envisaged where the fact that charges have been brought against the witness may have that effect. The application of the rule to outstanding charges, as the Crown accepts, is really just based on common sense and every day experience. No-one should now be in any doubt that the disclosure rule applies to them, or as to the reasons why this is so.

LORD WALKER:

30. I am in full agreement with the judgment of Lord Rodger. For the reasons that he gives I would dismiss this appeal.

LORD BROWN:

31. I agree with the judgment of Lord Rodger and, for the reasons that he gives, I too would dismiss this appeal.

LORD KERR:

32. I agree with the judgment of Lord Rodger and, for the reasons that he gives, I too would dismiss this appeal.

IN THE HIGH COURT OF JUSTICE

QUEEN'S BENCH DIVISION.
B e f o r e :
THE HONOURABLE MR JUSTICE TUGENDHAT.

Mr Choudhury (instructed by Taylor Vinters) for the
Claimant,
Mr Bowsher QC and Mr Palmer (instructed by Treasury
Solicitors) for the Defendant,
Hearing dates: 13 January 2010.

Mr Justice Tugendhat.

1. The Claimant, B2Net Limited, applies for an interim order to prevent the Defendant from continuing with a procurement exercise leading to the award of framework agreements for the provision of IT goods and services to government. It does so on the basis of a challenge to a single question contained in the Defendant's Pre-Qualification Questionnaire (PQQ), the responses to which determined which suppliers would be invited to tender in respect of a framework agreement.

2. The Defendant is the executive agency within HM Treasury tasked with providing certain services relating to procurement for the public sector. In particular, the Defendant facilitates framework agreements for a variety of products and services.

3. The Claimant contends that a procurement exercise being conducted by the Defendant is in breach of the requirements of the *Public Contracts Regulations 2006* ("the Regulations"), the relevant EC Directive (The Regulations implement Directive 2004/18/EC on the Co-ordination of procedures for the award of public works contracts) and general EC principles of non-discrimination and transparency. The breach means that the Claimant is excluded from proceeding to the tender stage of the procurement exercise despite having scored the maximum available marks in almost all categories in the PQQ. An interim order to suspend the exercise is one that is provided for by reg 47(8) where a breach of the duties owed to a person such as the claimant is alleged.

THE TEST TO BE APPLIED:

4. Mr Choudhury for the Claimant submits that the considerations governing an application for interim relief under reg 47(8)(a) are so similar to those which arise in an ordinary application for an interim injunction (see *American Cyanamid*) that it is appropriate to apply the same principles in determining whether such relief is appropriate: *Lettings International Ltd v London Borough of Newham* [2007] EWCA Civ 1522 at para 12. Accordingly, the Court must consider the following questions:
a. Is there a serious issue to be tried? If so,
b. Would damages be an adequate remedy; and
c. Does the balance of convenience favour maintaining the status quo?

5. Mr Bowsher QC for the Defendant does not dissent from this submission. But the position of the Defendant in these proceedings is not easily comparable to that of a defendant against whom interim relief is sought in private law proceedings. Mr

Bowsher submits that the true nature of the applicable principles is better derived from the recent statement of Lord Hoffmann in *National Commercial Bank Jamaica Ltd v Olint Corpn Ltd* [2009] UKPC 16; [2009] 1 WLR 1405:

> "16 … It is often said that the purpose of an interlocutory injunction is to preserve the status quo, but it is of course impossible to stop the world pending trial. The court may order a defendant to do something or not to do something else, but such restrictions on the defendant's freedom of action will have consequences, for him and for others, which a court has to take into account. The purpose of such an injunction is to improve the chances of the court being able to do justice after a determination of the merits at the trial. At the interlocutory stage, the court must therefore assess whether granting or withholding an injunction is more likely to produce a just result. As the House of Lords pointed out in *American Cyanamid Co v Ethicon Ltd* [1975] AC 396, that means that if damages will be an adequate remedy for the plaintiff, there are no grounds for interference with the defendant's freedom of action by the grant of an injunction. Likewise, if there is a serious issue to be tried and the plaintiff could be prejudiced by the acts or omissions of the defendant pending trial and the cross-undertaking in damages would provide the defendant with an adequate remedy if it turns out that his freedom of action should not have been restrained, then an injunction should ordinarily be granted.
>
> 17. In practice, however, it is often hard to tell whether either damages or the cross-undertaking will be an adequate remedy and the court has to engage in trying to predict whether granting or withholding an injunction is more or less likely to cause irremediable prejudice (and to what extent) if it turns out that the injunction should not have been granted or withheld, as the case may be. The basic principle is that the court should take whichever course seems likely to cause the least irremediable prejudice to one party or the other. This is an assessment in which, as Lord Diplock said in the *American Cyanamid* case [1975] AC 396, 408:
>
> > "It would be unwise to attempt even to list all the various matters which may need to be taken into consideration in deciding where the balance lies, let alone to suggest the relative weight to be attached to them."
>
> 18. Among the matters which the court may take into account are the prejudice which the plaintiff may suffer if no injunction is granted or the defendant may suffer if it is; the likelihood of such prejudice actually occurring; the extent to which it may be compensated by an award of damages or enforcement of the cross-undertaking; the likelihood of either party being able to satisfy such an award; and the likelihood that the injunction will turn out to have been wrongly granted or withheld, that is to say, the court's opinion of the relative strength of the parties' cases.
>
> 19. There is however no reason to suppose that in stating these principles, Lord Diplock was intending to confine them to injunctions which could be described as prohibitory rather than mandatory… What is required in each case is to examine what on the particular facts of the case the consequences of granting or withholding of the injunction is likely to be. If it appears that the injunction is likely to cause

irremediable prejudice to the defendant, a court may be reluctant to grant it unless satisfied that the chances that it will turn out to have been wrongly granted are low; that is to say, that the court will feel, … "a high degree of assurance that at the trial it will appear that at the trial the injunction was rightly granted."

6.   In his submissions Mr Bowsher addressed first the question whether an interim injunction should be made simply on the basis that damages would be a more than adequate remedy for the Claimant and, on the other hand, serious losses will be suffered by a range of other parties and there can be little certainty that all these losses will be adequately met. It not easy to identify who might be the other parties who might suffer if an interim injunction were granted and the Claimant failed at trial. There is evidence that the Defendant would suffer financially from the grant of an interim injunction, but Mr Bowsher does not advance that as a consideration (as a defendant in private law proceedings would). Rather, the losers will be: others who have submitted applications in competition with the Claimant; the public authorities for whose benefit the exercise is being held, and ultimately the public at large, as taxpayers and recipients of the services of the public authorities.

7. While bearing in mind that that is the main point advanced by Mr Bowsher, I shall (after first reciting the background) consider the issues in the order submitted by Mr Choudhury.

BACKGROUND:

8. The Claimant is an IT storage company. It provides companies with both hardware and software to improve IT performance, management and storage.

9. On 28 July 2009, by an OJEU notice, the Defendant commenced the competitive public procurement to establish a framework agreement for the provision of IT goods and services. The envisaged number of successful operators was 15 at that time. The framework agreement comprises three lots:

a. Lot 1 – Desktop hardware;

b. Lot 2 – IT infrastructure hardware; and

c. Lot 3 – Specialist channel partners for software.

10.   Responses to the PQQ in respect of the procurement were required to be submitted by 28 August 2009.

11.   The PQQ was in three sections:

A General Capability – 5 questions with an overall weighting of 20%;

B Lot Specific Capability - 20 questions with an overall weighting of 50%; and

C Previous Experience - 6 questions with an overall weighting of 30%. Of that 30%, the impugned question 6 Breadth of Experience had an overall weighting of 7.5%.

12.   The Claimant submitted its response to the PQQ by the required date with a view to being selected for an invitation to tender ("ITT") in respect of Lot 2 of the framework agreement.

13.   On 1 October 2009, the Claimant was informed that the Defendant's evaluation of the responses to the PQQ had been completed and that the Claimant had been unsuccessful.

14.   From the information provided by the Defendant in a revised debrief document dated 3 December, it was apparent that:

i) 20 out of 84 competing suppliers had been successful at the PQQ stage and would be invited to tender for Lot 2;

ii) The range of scores of successful suppliers was 935.71 to 970.00 out of an available maximum of 970.00. The original maximum of 1000 points was reduced to 970 following the withdrawal by the Defendant of

one of the questions on the PQQ as a result of a number of challenges brought by bidders;

iii) The Claimant's score was 922 out of 970. It had therefore failed to be shortlisted by a margin of only 13.71 points (1.4%):

iv) The Claimant had scored the maximum available points in respect of all but two of the questions set out in the PQQ. These two were:

a) Question A5 (Quality Management System) - The Claimant scored 12 out of 30 for this question because its ISO 9001 accreditation was still pending. The Claimant does not take issue with this criterion or the score awarded;

b) Question L2 C6 (Breadth of Experience) – The Claimant scored 45 out of 75 for this question.

15.   Section C of the PQQ, entitled "Previous Experience" required the Claimant to provide five example contracts from the last three years relevant to Lot 2. Points were awarded in respect of each such contract based on its relevance to the Lot being applied for. The Claimant scored the maximum in respect of each of the contract examples submitted.

16.   In the same section of the PQQ, question L2 C6 was in the following terms:
"[L1to3 C6 Breadth of Experience]
Separately to the above, marks will be awarded for demonstrating a breadth of experience across the full range of products and services relevant to each Lot. If all 5 examples provided for each Lot are relevant, they will be considered together and an additional mark awarded as below:
1. Each of the 5 examples was awarded directly to the bidding organisation
2. Each of the 5 examples are drawn from different customers …
3. The 5 examples overall demonstrate capability across the full range of products and services relevant to each Lot.
The marking scheme is as follows:
0 Not all of the 5 examples are relevant or neither criteria are met
1 All 5 examples are relevant and one of the criteria is met
3 All 5 examples are relevant and two of the criteria are met
5 All 5 examples are relevant and three of the criteria are met
Please note, no response to this question is required." [Emphasis Added by the Claimant]

17.   This was the only question in the PQQ deemed to be optional. I have been unable to see any significance in this point.

18.   The questions asked by the Defendant included:
"Confirm the contract was placed direct with your organisation or name the prime contractor concerned"

19.   The answer given by the Claimant in respect of four of the five examples was in two parts which I have numbered:
"[1] Prime Contractor DSGI
[2] All client engagement has been managed directly between B2Net and the customer with DSGI facilitating the purchase through the existing catalist framework. The contract was held between B2Net and the end user".

20.   I shall refer to part [2] of the answers as the Explanation.

21.   Upon receipt of the Defendant's notification that it had been unsuccessful, the Claimant sought further information as to the reason for its failure to score the maximum points under section C of the PQQ.

22.   On 13 October 2009, the Defendant responded as follows:

"B2Net were awarded 3 marks for L2C6 as all 5 examples provided were relevant and two of the criteria were met. Four of the contract examples were not awarded directly to the bidding organisation. The response provided states that DSGI was the prime contractor in each of those examples. Therefore point 1 above [i.e. the criterion that each of 5 examples was awarded directly to the bidding organisation] was not met. We are satisfied that the scoring of this question is correct and consistent with the instructions provided within the PQQ."

23. On 14 October the Claimant replied stating that the writer had anticipated that this would be the area where the Claimant failed to score full marks. The letter continued:

"During the PQQ stage … the following question was asked [by another bidder] and answered [by the Defendant in a form communicated to the other bidders]

Q72 With regards to the scoring scheme for example contracts, where a contract was placed directly, owned and driven by the reseller, but a 3rd party was used purely as an invoicing mechanism, will this be scored in line with the 2 point criteria rather than the 5 point criteria…[?]

A72 Such an arrangement would not preclude the Example Contract from scoring 5 points so long as the contract was between the customer and the bidding organisation and not with the 3rd party organisation supplying the invoicing mechanism ….

With all the examples offered by B2Net the customer's contract was always delivered by B2Net. The fact that most public sector organisations require to use OGC as a procurement framework means the requirement for a 'direct contract' is very difficult to provide simply due to the frameworks already in place.

DSGI are a partner to use simply as a transactional partner in these instances and are literally only an invoicing mechanism to satisfy procurement rules.

The contract, the delivery and the ongoing support of the solutions we deploy are entirely between B2Net and the end user customer".

24. That is the gist of the challenge by the Claimant in these proceedings. The Claimant says that the omission to give full marks to a bidder who failed to give an example in which he had been a prime contractor was in breach of the Regulations.

25. The Defendant responded on 23 October 2009 confirming that it was satisfied that the scoring was correct.

26. On 12 November 2009, the Defendant wrote to all bidders stating that question A4 "Growth of Business" in the PQQ had been challenged on the grounds of validity and had been removed. Revised results were subsequently issued to all suppliers on 10 December 2009. In the course of revising the results, the Defendant also increased the number of successful suppliers that would proceed to the ITT stage. In the course of this revision, the Defendant drew the line between successful and unsuccessful bidders for Lot 2 immediately above the Claimant's score. This meant that it was not until 10 December 2009 that final confirmation was received as to the successful bidders.

27. On 3 December 2009, the Claimant's solicitors gave notice to the Defendant that the scores awarded to the Claimant in the PQQ would be subject to challenge and invited the Defendant to defer any further decision-making in the meantime. The grounds of the Claimant's challenge were set out in a further letter dated 16 December 2009. In particular, it was asserted that question L2C6 was invalid and the grounds for that assertion were set out. The Defendant was once again invited to revisit the scoring process and to confirm that the Claimant would be invited to tender, failing which the Claimant would make a formal challenge under the 2006 Regulations.

28. By a letter dated 17 December 2009, the Defendant responded to the Claimant's notice by stating that it did not understand why question L2C6 is considered to be invalid and seeking further explanation from the Claimant. The Defendant further confirmed that the procurement timetable had been adjusted in that invitations to submit tenders were sent to selected suppliers on 11 December 2009 and that the deadline for the receipt of tenders is now 26 January 2010. However, although tenders would now be received about 10 weeks after the original deadline of 11 November 2009, the Defendant only moved the contract issue date by 4 weeks. The marketing launch date of 1 March 2010 remains the same.

29. The Claimant's Application Notice was issued on 23 December 2009 with notice of hearing on 7 January 2010. The first response to the Application Notice was not received until 4 January 2010. The Defendant's evidence in response was served shortly before 1.00pm on 6 January 2010.

30. It is common ground that the provisions under which Q72 was asked and answered would have permitted the Claimant to ask a corresponding question about what it states was its relationship between DSGI and the customers in the examples which it gave. The Claimant did not take this opportunity, but raised the issue for the first time as set out above.

THE LAW APPLICABLE TO THE PROCUREMENT PROCESS:

31. There is no dispute as to the requirements of the Regulations. They are summarised by Mr Choudhury as follows.

32. The Defendant is required to conduct procurement exercises in accordance with the Regulations, the Directive 2004/18/EC and general principles of EC Law.

33. In particular, the Defendant is required to treat the Claimant equally with other economic operators and in a non-discriminatory way; and to act in a transparent way: Reg. 4(3)

34. In respect of any procurement conducted in accordance with the restricted procedure set out in reg 16 of the Regulations, the Defendant is required to make its evaluation of economic operators in accordance with regs 23, 24, 25 and 26, and may exclude an economic operator from those economic operators from which it will make the selection of economic operators to be invited to tender only if the economic operator:

    i) may be treated as ineligible to tender on a ground specified in Regulation 23; or

    ii) fails to satisfy the minimum standards required of economic operators by the Defendant of:

        a) Economic and financial standing; or

        b) Technical or professional ability: Reg 16(7)

35. In assessing whether an economic operator meets any such minimum standards of technical or professional ability, the Defendant may have regard to any means

listed in reg 25(2) of the Regulations according to the purpose, nature, quantity or importance of the contract. Those means include, in the case of a public services contract, a public works contract or a public supply contract requiring the siting or installation of work, the economic operator's technical ability, taking into account in particular that economic operator's skills, efficiency, experience and reliability.

36. The Regulations do not mention the economic operator's status, i.e. whether as a contractor, sub-contractor or as part of a consortium, in acquiring or otherwise evidencing such technical or professional ability.

37. The Defendant is also entitled to limit the number of economic operators which it intends to invite to tender, provided that the contract notice specifies the objective and non-discriminatory criteria to be applied in order to limit the number of such operators: Reg 16(9).

38. Reg 47 provides that breach of the Regulations is actionable by any economic operator which, in consequence, suffers or risks suffering loss or damage and those proceedings shall be brought in the High Court.

39. In *Lion Apparel Systems Limited v Firebuy Limited* [2007] EWHC 2179 (Ch), [2008] EuLR 191 Morgan J set out the legal principles applicable to the procurement processes such as the one here in question. These included:

> "35. The court must carry out its review with the appropriate degree of scrutiny to ensure that the above principles for public procurement have been complied with, that the facts relied upon by the Authority are correct and that there is no manifest error of assessment or misuse of power.
> 36. If the Authority has not complied with its obligations as to equality, transparency or objectivity, then there is no scope for the Authority to have a "margin of appreciation" as to the extent to which it will, or will not, comply with its obligations.
> 37. In relation to matters of judgment, or assessment, the Authority does have a margin of appreciation so that the court should only disturb the Authority's decision where it has committed a "manifest error".
> 38. When referring to "manifest" error, the word "manifest" does not require any exaggerated description of obviousness. A case of "manifest error" is a case where an error has clearly been made.
> 39. I take the above principles from the decision of the Supreme Court of Ireland in *Siac Construction v Mayo County Council* [2003] EuLR 1, and the decision of the Court of First Instance in *Evropaiki Dynamiki v Commission* 12th July 2007 at [89]".

40. It follows that I have to consider whether the Claimant has raised a serious issue to be tried as to whether the Defendant has breached any obligation under the Regulations or, in assessing the Claimant's response to PQQ, made a manifest error.

IS THERE AN ISSUE TO BE TRIED?

41. The most important submissions advanced are in my view the following.

42. Mr Choudhury's first submission is that in the first 5 questions in Section C one of the factors identified in PQQ was whether the contract given as an example was placed directly with the applying organisation. The Claimant got full marks on that section, notwithstanding that in four out of its five examples it was a sub-contractor. Mr Choudhury submits that this raises issues of consistency and

transparency: why should the Claimant not have got full marks for the sixth question?

43. Mr Bowsher submits that there is no direct comparison between the two sets of questions. The first five questions are so framed that it is clear from PQQ that it is possible for a sub-contractor to score full marks on those questions.

44. It appears to me that as a matter of construction the Claimant's case is weak on this point.

45. Next Mr Choudhury submitted that there is no rational explanation for marking down a sub-contractor, because a sub-contractor may have, and in this case the Explanation shows that the Claimant did have, all the relevant experience required. It is said that the Claimant had actually done the work under the contracts given as examples. In such cases the prime contractor will have less experience than the sub-contractor, and yet the PQQ system of marking gives the prime contractor a preference.

46. There has been no evidence before me, in these interim proceedings, of the meaning of Q72 and A72 (set out in the Claimant's letter of 14 October 2009), and no investigation into the facts of the relationship between the Claimant and DSGI and the customers of DSGI and/or the Claimant in the examples given. It is in issue whether the Explanation that the Claimant gave is either (a) comparable to the facts described in Q72, or (b) such that the Claimant's experience is to be regarded as in all respects similar to that of a prime contractor. These will, or might, be issues for consideration at trial.

47. Mr Bowsher submits that if the other questions in PQQ are considered, a prime contractor who has employed a sub-contractor will be identified by his responses to other questions, and so will lose marks at that point, and not, in the end, be accorded a preference over sub-contractors merely as a result of his status as a prime contractor.

48. Mr Bowsher submits that the relative marking of prime contractors and sub-contractors is a matter within the margin of appreciation allowed to the Defendant.

49. Further, Mr Bowsher submits that it would not have been open to the Defendant to give the Claimant marks for the Explanation, since there had been no publication to others that that might be done. It might have been otherwise if the Claimant had asked, and been given a public answer, to a question corresponding to Q72. But in any event, the Explanation raises, or might raise further questions.

50. There are a number of other ways in which Mr Choudhury advances this, or a similar point. He submits that the Defendant has given preference to form over substance, and that there is no satisfactory explanation for the marking down of sub-contractors given in the evidence. And in so far as any explanation is given in the evidence, then it raises an issue of transparency: the explanation should have been given in PQQ.

51. This is a point on which the court is not well placed to form a view at this stage of the proceedings. My preliminary view is that there is likely to be a material and objectively justifiable difference between a prime and a sub contractor from the point of view of the Defendant. My preliminary view is that the letter of 14 October by the Claimant is unconvincing in seeking to assimilate the two, even where the sub-contractor has in effect done all the work. In so far as I am able to form a view of the strength of this point, it appears to me that on this point too the Claimant's case is weak.

52. Given the approach of Mr Bowsher, I am prepared to assume that the Claimant may have raised a serious issue to be tried, but I cannot say that I consider it to be a strong case. On this basis I do not need to consider the merits of the claim further.

ADEQUACY OF DAMAGES FOR THE CLAIMANT:

53. In his first witness statement for the Claimant Mr Thompson stated that if an injunction is not granted the Claimant will not be able to participate in the process at all and that there would be no prospect whatsoever of being a party to the framework agreement. He goes on to say that "In these circumstances, damages would clearly be a wholly inadequate remedy". He gives no explanation for this conclusion.

54. This was pointed out for the Defendant by Mr Cliffe. He stated that the Claimant was not on the existing framework, but had done business as a sub-contractor to a prime contractor who was on the existing framework. There was no evidence to explain why it should not continue to do so.

55. Mr Thompson made a second witness statement. He said that being on the new framework would enhance the Claimant's reputation. By this I understand him to mean that the fact that the Claimant was on the framework would give rise to a chance of it obtaining work (whether under or outside the framework) which it would not have if it fails in these proceedings. It would also increase its margin, in that there would be none for the prime contractor. He estimated the increase would be a percentage which he specified. Moreover, DSGI, which was the prime contractor through which it had dealt under the existing framework, was not amongst the 20 selected to tender for the new framework. Accordingly, the Claimant would have to deal through a substitute prime contractor, quite possibly on less favourable terms as to margin.

56. Mr Bowsher submits that (assuming no interim relief is granted, but the Claimant succeeds on liability), at the time when this claim would come to the assessment of damages, the new framework will have been in operation for some time, and there will be data from which margins and other relevant figures can be found for the purposes of assessing damages. He accepts that the damages may not be as good a remedy as an injunction, but submits that they will be adequate.

57. As to the law, Mr Choudhury submits that the court should take care not to set too high a standard, since that would be to deprive claimants of the effective remedy which reg 47 is intended to provide.

58. In response to that Mr Bowsher notes that a higher test for interim relief has been applied in the Court of First Instance in Case T-511/08R, *Unity OSG FZE v Council of the European Union* (Order of 23 January 2009), and so that there a test which applies any similar or lower threshold would not be unlawful. In that case the court said:

> "It must be noted that the urgency of an application for interim measures must be assessed in relation to the necessity for an interim order in order to prevent serious and irreparable damage to the party applying for those measures. It is for that party to prove that it cannot wait for the outcome of the main proceedings without suffering damage of that kind…. "

59. Applying the test in *Cyanamid,* and accepting the evidence of Mr Thompson, there is nothing upon which I should find that damages would not be an adequate remedy in this case. On the contrary, the evidence suggests that damages would be an adequate remedy, and more readily capable of calculation

than many claims for damages for loss of business that come before the courts. The longer any assessment of damages is deferred into the term of the new framework, the more evidence there will be.

60. I turn to consider the possible injustice if an interim injunction is granted, but the Claimant fails to establish his case at trial.

61. Mr Bowsher submits that an interim injunction would cause significant losses and other prejudice to a large number of entities, both private and public. Given the extensive range of public purchasers that are expected to use this framework (not least because they used the framework agreement which this replaces), it has not been possible to gather comprehensive evidence regarding the impact of delay in implementation of this framework. But there is some evidence from Mr Cliffe. He states that the Defendant facilitates the buying process in a vast and highly complex marketplace providing access to over 500,000 products and services through more than 600 suppliers. The customer base spans the biggest central government departments, NHS Trusts and local councils, through to the smallest schools. I understand that evidence to relate to its activities generally. In relation to the existing framework (due to expire on 30 April 2010) there were orders from customers of £353m between April and September 2009, which he states represents savings of some £23.9m to the UK public sector. There might also be losses suffered by other bidders in respect of the delay to the current procedure that would follow from the grant of an interim injunction.

62. It seems to me very unclear how any losses that might be suffered by public sector buyers or by other bidders could be advanced in a claim on any cross-undertaking in damages. But that does not mean that there would be no damage done by the grant of an injunction. The disruption must inevitably be damaging, or so it seems to me. The remedy under a cross-undertaking, however framed, does not appear to me to be one that would be adequate to prevent injustice. That will not of itself preclude the grant of an injunction, for which the Regulations make specific provision. But it is a factor to be considered.

63. There was some debate between the parties as to the time for which any suspension of the procedure would be likely to last, when a trial of this action might take place, and whether or how the existing framework could be extended to cover that period. I do not need to consider this point in detail. It is difficult to predict what the issues might be in the trial. It might be tried substantially on the documents before me, or it might give rise to complicated disclosure and factual issues. It is impossible to predict when a trial might take place or the period for which the suspension would be required.

64. If I had formed the view that the Claimant's case on the merits was a strong one, then that might have weighed in the balance in its favour. But that is not this case.

65. Accordingly I dismiss this application on the ground that damages would be an adequate remedy for the Claimant, but not for the numerous other parties who would be affected by the suspension of the procedure which the Claimant seeks.
OTHER POINTS:

66. Mr Bowsher advanced a number of other points on which I can state my conclusions very briefly. I would not have refused an injunction (if it were otherwise just to grant it) on the basis that the Claimant could not give an adequate cross-undertaking in damages. In this case that would not be a sufficiently significant factor.

67. Mr Bowsher takes no point on the three month limitation period (explaining that the judgment of the ECJ on this point in *Case C-406/08*, *Uniplex (UK) Ltd v NHS BSA* is expected to appear very shortly). Mr Bowsher does rely on the delay that has occurred. The loss of marks to those with experience as sub-contractors was evident when the PQQ was published. The Claimant did not raise the point until October, and then delayed commencing proceedings until 23 December.

68. There is force in this point. Had I been otherwise undecided, this would have weighed significantly against the Claimant in my judgment. See especially *Jobsin Internet Services v Department of Health* [2001] EWCA Civ 1241, [2001] EuLR 685 paragraphs 33 & 38

CONCLUSION:

69. For these reasons I dismiss this application.

IN HER MAJESTY'S COURT OF APPEAL
IN NORTHERN IRELAND.
_____

APPEAL BY WAY OF CASE STATED
FROM A DECISION OF THE FAIR
EMPLOYMENT TRIBUNAL IN
ACCORDANCE WITH ARTICLE 90 OF
THE FAIR EMPLOYMENT AND
TREATMENT (NORTHERN IRELAND)
ORDER 1998 AND ORDER 61 OF THE
RULES OF THE SUPREME COURT
(NORTHERN IRELAND) 1980.

_____

[1]   This is an appeal by way of case stated from the Fair Employment Tribunal ("the Tribunal") delivered on 14 April 2008. The respondents/appellants were represented by Ms Noel McGrenera QC and Mr Jonathan Dunlop while the applicant/respondent is a litigant in person.

Background facts:

[2]   In 1999 the respondent was a serving officer in the Royal Ulster Constabulary, now the Police Service of Northern Ireland and on 6 August 1999 and Internal Force Message was circulated amongst serving officers inviting applications for officers to be seconded to Kosovo for a 12 month period of deployment commencing in mid-September 1999. The request for the deployment, which emanated from the United Nations and was co-ordinated by the Home Office, was for a total of 60 officers comprising a superintendent as contingent commander, two inspectors, eight sergeants and forty nine constables with two sergeants and four constables to act as reserves.

[3]   One hundred and six officers applied in response to the invitation and a "paper sift" was carried out in August 1999 by the first named appellant. A total of 71 officers, including the respondent, were selected following the paper sift.

[4]   The next stage of the selection process was a training course that was to be held at various police venues in Northern Ireland between 27 September and 15 October 1999 to be followed by a week's attendance at the training centre of the Garda Siochana at Templemore, County Tipperary.

[5]   The second named appellant, who was then serving as the Deputy Regional Head of CID for the Belfast Region, was selected by the Chief Constable to be the Superintendent in charge of the contingent and on 14 October 1999 the second named appellant decided that the respondent was to be included in the reserve list rather than amongst the 60 officers to be initially deployed. On 14 October 1999 the second named defendant communicated his decision to the officers concerned informing them that only 60 were required by the United Nations although the Chief Constable had confirmed that he was quite prepared to permit all 68 officers to be deployed, including the eight placed on the reserve list. The second named defendant explained that no stigma should be attached to an officer placed on the reserve list and that all officers should complete the course since he had no doubt that the reserves would be deployed

before the year ended and, if not, they would be included in the next deployment. The respondent was clearly dissatisfied with being placed on the reserve list and emphasised to the second named appellant his desire to be included among those initially deployed. When the second named appellant reminded him that he had a number of outstanding cases including a number of assaults on the police which were invariably contentious he insisted that they could all be "sorted out" leaving him free to be included. On 15 October 1999 the respondent again spoke to the second named appellant at the canteen at Garnerville training facility. And there was a further discussion about his selection for the reserve list.

[6]   On 4 November 1999 the respondent submitted an application to the Tribunal making complaints of victimisation, sexual and religious discrimination. The hearing was conducted before the Tribunal between 8-12, 15-19 and 30-31 October 2007, the 15 and 16 November 2007 and 13 December 2007. On 14 April 2008 the Tribunal delivered its decision unanimously dismissing the respondent's claims of victimisation and sexual discrimination but upholding the claim of direct religious discrimination.

[7]   On 21 May 2008 the appellants submitted a requisition to the Tribunal to state a case raising eight questions for the opinion of this court. On 1 October 2008 the Tribunal stated a case helpfully reducing the questions for the opinion of this court to a total of two. These are:

"(*i*) Whether the Tribunal, on the facts proved or admitted was correct in law in deciding the appellants had not discharged the burden of proof, pursuant to Article 38A of the Fair Employment and Treatment (Northern Ireland) Order 1998?

(*ii*) Whether the Tribunal's decision, on the facts proved or admitted, was a decision which no reasonable Tribunal could have reached and was perverse in law?"

The evidence before the Tribunal:

[8]   A wide range of issues were canvassed before the Tribunal during a hearing which lasted some 15 days and produced a judgment running to some 44 pages. That judgment was highly critical of the procedure adopted by the second named appellant for the purpose of selecting those who were to be included in the initial deployment and referred to it as having been carried out "in a somewhat informal/ad hoc way" with "no documentation/record properly kept" to demonstrate the basis upon which the assessments had been made. The second named defendant maintained that his decision to place officers on the reserve list had been based upon an assessment of various specific criteria including application scores, sick records, performance on the training course as described by other supervisors, complaints against officers and outstanding court cases. The Tribunal recorded that, in such circumstances, it would have expected to be furnished with proper detailed document/records identifying specific candidates and clearly and transparently recording the assessment of each such candidate against the relevant criteria. The second named appellant was unable to give detailed evidence of the basis upon which the performance of candidates during the training course had been assessed explaining that it came down to a matter of judgment on his part based on his experience. He said that if no adverse comment had been made about any particular candidate he assessed that candidates performance as "good" and did not further investigate the matter.

[9]   The Tribunal recorded that the crucial factor relied upon by the second named appellant as the basis for his decision to include the respondent in the reserve list had been a specific adverse comment that the second named appellant alleged had been made about the respondent's performance during the training course. The comment was that the respondent had been over enthusiastic in relation to the use of handcuffs. The

266

Tribunal described the second named appellant as being "extremely vague" about this comment, being unable to remember the circumstances under which it had been made, and by whom it had been made although he believed that it had been made by one of the trainers and relayed by one of the training inspectors. During his evidence he expressed the view that it had probably been reported by him by Inspector Douglas. The second named appellant explained that, as a result of hearing this comment, he had concerns about the respondent's suitability in the volatile environment of Kosovo and that, as a result, he made enquiries of the Personnel Department in order to discover whether any allegations/complaints had been made by members of the public against the respondent. He said that he was informed by the Personnel Department that there had been complaints/allegations against the respondent by members of the public which related to alleged assaults and incivility. He agreed that he had not obtained any records or other details when making his enquiry. No such enquiries were raised with the Personnel Department about any other participant in the training course and the second named appellant maintained that such action was unnecessary in the absence of a similar adverse comment.

[10]   The Tribunal rejected the second named appellant's evidence that he had received an adverse comment about the performance of the respondent during the training course for the following reasons:

(*i*) There was no written record of receiving the comment.

(*ii*) The second named appellant had not included any specific reference to the comment in either his contemporaneous journal or witness statements.

(*iii*) Despite the significance of the comment it had not been mentioned by the second named appellant to the respondent on either 14 or 15 October at times when the respondent had obviously been very anxious to learn as much as possible about the reason for being placed on the reserve list.

(*iv*) The said comment had not been referred to during the subsequent grievance procedure brought by the respondent.

(*v*) The second named appellant had been extremely vague about this aspect of his evidence.

(*vi*) Despite expressing the view that the comment had probably been made by Inspector Douglas, the second named appellant had not called that officer as a witness. In such circumstances, the Tribunal came to the conclusion that Inspector Douglas' evidence would not have supported the second named appellant in accordance with the decision in Lynch v Ministry of Defence [1983] NI 216.

[11]   Having rejected the second named appellant's evidence about the alleged adverse comment on the respondent's performance in the training course. The Tribunal concluded that the respondent had established facts from which the Tribunal could conclude, in the absence of an adequate explanation, that the appellants had committed an act of discrimination against the respondent on the grounds of religious belief. In reaching those conclusions the Tribunal relied upon the provisions of Article 38A of the Fair Employment and Treatment (Northern Ireland) Order 1998 (the 1998 Order) and the jurisprudence relating to the interpretation thereof including Igen v Wong [2005] IRLR 258, Laing v Manchester City Council [2006] IRLR 748, Madarassy v Nomura International Plc [2007] IRLR 246, McDonagh and Others v Samuel Tom T/as The Royal Hotel Dungannon [2007] NICA 3 and Arthur v Northern Ireland Housing Executive and SHL (UK) Ltd [2007] NICA 25.

The relevant law:

[12]   Part III of the 1998 Order prohibits discrimination in the field of employment and Article provides as follows:

"(1) In this order 'discrimination' means –

(*a*) Discrimination on the ground of religious belief or political opinion;

(2) The person discriminates against another person on the ground of religious belief or political opinion in any circumstances relevant for the purposes of this order if –

(*a*) On either of those grounds he treats that other person less favourably than he treats or would treat other persons;"

Article 38A of the 1998 Order which relates to the burden of proof provides as follows:

"Where on the hearing of a complaint under Article 38, the complainant proved the facts from which the Tribunal court apart from this Article, conclude in the absence of an adequate explanation that the respondent –

(*a*) committed an act of unlawful discrimination … against the complainant; or

(*b*) is by virtue of Article 35 or 36 to be treated as having committed such an act of discrimination … against the complainant, the Tribunal shall uphold the complaint unless the respondent proves that he did not commit or, as the case may be, is not to be treated as having committed that act."

Conclusions:

[13]   The appellant's advisors criticised the Tribunal's rejection of the second named appellant's evidence relating to the adverse comment upon a number of grounds. For example, they submitted that it was hardly surprising that the second named appellant had not recorded the comment in the context of the Tribunal's finding that he had generally carried out the selection exercise in an "informal/ad hoc way" without properly keeping any documents or records. They also emphasised the fact that neither the respondent nor the Tribunal had ever directly suggested to the second named appellant that he had fabricated his evidence about the comment and, indeed, that the fact that he had not done so was supported to some extent by the reference at paragraph 5 of the written statement by Chief Superintendent Wilson to the fact that the second named appellant had provided course performance as one of the reasons for placing the respondent on the reserve list. The witness statement made by Chief Inspector, as he then was, Wilson was admitted before the Tribunal as hearsay evidence on behalf of the claimant – see paragraph 3.18 of the Tribunal's decision. They further submitted that the fact that the second named appellant had spoken to the inspectors and trainers about conduct on the training course would have been clear from the second named appellant witness statement and journal entry. The plaintiff's own witness statement confirmed that the second named appellant had told him that he had spoken to and taken into account the comments made by the inspectors responsible for the training course specifically recording that:

"I asked Superintendent Middlemiss did the directing staff from COT trainers say anything about my performance during training. He replied, 'Yes, it is because of comments made and your Courts list that you are on the reserve list'."

The appellant's advisors also drew the attention of the court to the fact that, apart from the reference to the adverse comment, the Tribunal had been prepared to accept and relied upon every other key point in the second named appellant's evidence.

[14]   It is clear from the relevant authorities that the function of this court is limited when reviewing conclusions of facts reached by the Tribunal and that, provided there was some foundation in fact for any inference drawn by a Tribunal the appellate court should not interfere with the decision even though they themselves might have preferred a different inference. As Carswell LCJ, as the then was, observed in <u>Chief Constable of the Royal Ulster Constabulary and Assistant Chief Constable A H v Sergeant A</u> [2000] NI 261 at 273:

"[4] The Court of Appeal which is not conducting a rehearing as on an appeal, is confined to considering questions of law arising from the case.

[5] A Tribunal is entitled to draw its own inferences and reach its own conclusions, and however profoundly the appellate court may disagree with its view of the facts it will not upset its conclusions unless –

(*a*) there is no or no sufficient evidence to found them, which may occur when the inference or conclusion is based not on any facts but on speculation by the Tribunal (Fire Brigade Union v Fraser [1998] IRLR 697 at 699, per Lord Sutherland); or

(*b*) the primary facts do not justify the inference or conclusions drawn but lead irresistibly to the opposite conclusion, so that the conclusion reached may be regarded as perverse; Edwards (Inspector of Taxes) v Bairstow [1956] AC 14, per Viscount Simmons at 29 and Lord Radcliffe at 36."

[15]   However, this court would wish to emphasis the need for a Tribunal engaged in determining this type of case to keep in mind the fact that the claim is founded upon allegation of religious discrimination. The need to retain such a focus is particularly important when applying the provisions of Article 38A of the 1998 Order. In both the decision and the case stated the Tribunal recorded that it had taken into account the fact that both Protestants and Catholics were selected for deployment, that both Protestants and Catholics were included in the reserve list and that the second named appellant, who was a Protestant, had previously been married to a Catholic and that his children and grandchildren were Catholic. However, in this context, another finding of fact by the Tribunal which was in our view fundamental was that, prior to the selection process for the reserve list, the second named appellant did not know the respondent – see paragraph 6.4 of the Tribunal's decision and paragraph 3.1(*7*) of the case stated. Neither the decision nor the case stated contains any reference as to whether, and if so how, the Tribunal gave specific consideration to the basis upon which this complete lack of prior knowledge of the respondent by the second named appellant could be reconciled with an inference of religious discrimination.

[16]   In Laing v Manchester City Council [2006] 1519, the case of alleged racial discrimination, Elias P said at paragraph 71:

"There still seems to be much confusion created by the decision in Igen [2005] ICR 931. What must be borne in mind by a Tribunal faced with a race claim is that ultimately the issue is whether or not the employer has committed and act of race discrimination. The shifting and the burden of proof simply recognises that there are problems of proof facing an employee which it would be very difficult to overcome if the employee had at all stages to satisfy the Tribunal on the balance of probabilities that certain treatment had been by reason of race."

In the Sergeant A case Carswell LCJ, as he then was, said at page 273:

"[3] Discrepancies in evidence, weaknesses in procedures, poor record-keeping, failure to follow established administrative processes or unsatisfactory explanations from an employer may all constitute material from which an inference of religious discrimination may legitimately be drawn. But Tribunals should be on their guard against the tendency to assume that every such matter points towards a conclusion of religious discrimination, especially where other evidence shows that such a conclusion is improbable on the facts."

[17]   In this case the Tribunal purported to follow the guidelines set out in Barton v Investec Henderson Crosthwaite Securities Limited [2003] IRLR 332 as approved in the Court of Appeal in England and Wales in Igen v Wong [2005] IRLR 258 and in this jurisdiction in McDonagh and Others v Samuel Tom T/as The Royal Hotel, Dungannon(2007) NICA 3. The approach that it adopted was first to consider in

isolation the second named appellant's evidence relating to the adverse comment and, having rejected that evidence, to conclude that the respondent had established facts from which the Tribunal could infer that the appellants had committed an act of discrimination against the respondent, namely, treating unfavourably by comparison with his Protestant comparators by consulting the records of public complaints held by Department B and doing so on the ground of his religion. In our view this was a flawed and over mechanistic approach as a result of which the Tribunal appears to have failed to give consideration to facts of fundamental importance namely that neither the respondent nor his religious persuasion had been known to the second named appellant prior to the selection exercise. At paragraph 4.4 of the original decision in the course of a careful analysis of relevant authorities the Tribunal included the following words from the decision of the Court of Appeal in England and Wales in <u>Madarassy v Nomuri International Plc</u> [2007] IRLR 246:

"The burden of proof does not shift to the employer simply on the claimant establishing a different in status (eg sex) and a difference in treatment. Those bear facts only indicate a possibility of discrimination. They are not without more, sufficient material from which a Tribunal 'could conclude' that, on the balance of probabilities, the respondent had committed an unlawful act of discrimination. 'Could conclude' in Section 63A(2) must mean that 'a reasonable Tribunal could properly conclude' from all the evidence before it. This would include evidence adduced by the claimant in support of the allegations of sex discrimination, such as evidence of a difference in status, difference in treatment and the reason for the differential treatment. It would also include evidence adduced by the respondent contesting the complaint. Subject only to the statutory 'absence of an adequate explanation' at this stage, the Tribunal needs to consider all the evidence relevant to the discrimination complaint, such as evidence as to whether the act complained of occurred at all, evidence as to the actual comparators relied on by the claimant to prove less favourable treatment, evidence as to whether the comparisons made by the complainant were of like with like as required by Section 5(3), and available evidence of the reasons for the differential treatment …. Although Section 63A(2) involves a two-stage analysis of the evidence, it does not expressly or impliedly prevent the Tribunal at the first stage from hearing, accepting or drawing inferences from evidence adduced by the respondent disputing and rebutting the claimant's evidence of discrimination. The respondent may adduce evidence at the first stage to show that the acts which are alleged to be discriminatory never happened; or that, if they did, they were not less favourable treatment of the claimant; or that the comparators chosen by the claimant or the situations with which comparisons are made are not truly like the claimant or the situation of the claimant; or that, even if there has been less favourable treatment of the claimant, it was not on the grounds of her sex or pregnancy (in this case religion). Such evidence from the respondent could, if accepted by the Tribunal, be relevant as showing that, contrary to the claimant's allegation of discrimination, there is nothing in the evidence from which the Tribunal could properly infer a prima facie case of discrimination on the prescribed ground."

The Tribunal also referred to the view of Elias J in <u>Laing</u>, quoted with approval by Campbell LJ in the <u>Arthur's</u> case, that it was obligatory for a Tribunal to go through the formal steps set out in <u>Igen</u> in each case. As Lord Nicholls observed in <u>Shamoon v Chief Constable of the Royal Ulster Constabulary</u> [2003] NI 174:

"Sometimes a less favourable treatment issued cannot be resolved without, at the same time, deciding the reason why issue. The two issues are intertwined."

[18]   In relation to the respondent's allegation of sex discrimination the Tribunal correctly applied the observations of Mummery LJ in <u>Madarassy</u> in holding that simply

proving unfavourable treatment and a different status, in that context sex, gave rise merely to a possibility of discrimination and was not sufficient to shift the burden of proof. The Tribunal recognised that a similar situation existed in relation to the respondent's claim for religious discrimination insofar as he had established unfavourable treatment and a difference of status, in this context religion, between himself and Constables R and B but again accepted that those facts alone would not have been sufficient to shift the burden of proof. However, the crucial difference for the Tribunal appears to have been its finding that the evidence of the second named appellant relating to the adverse comment had not been made. At paragraph 7.7 the Tribunal stated that it had no hesitation in concluding that the burden of proof had shifted as a consequence of this finding. In our view that was a flawed approached to the evidence. The evidence about the making of the adverse comment was the rationalisation put forward by the second named appellant for carrying out the enquiries with Department B. The Tribunal found not only that such enquiries had been made by the second name appellant but that such enquiries would have been reasonable and appropriate had the adverse comment been made. In the circumstances we consider that the proper approach for the Tribunal to have adopted would have been to consider that rationalisation in the context of the surrounding evidence and not in isolation in relation to the issue as to why the enquiries with B Department were made about an officer whose identity and religion had been completely unknown prior to and during the selection process. In such circumstances only one inference could reasonably have been drawn, namely, that the enquiries were stimulated by a comment of the nature described by the second named appellant rather than on the ground of the respondent's religion.

SHERIFFDOM OF LOTHIAN AND
BORDERS AT EDINBURGH.
Judgement of
Sheriff Kathrine EC Mackie
In causa
MR MARK ADAMS residing at 20a Binny
Park Broxburn EH52 6NP, PURSUER,
against
THE NATIONAL INSURANCE AND
GUARANTEE CORPORATION LIMITED
a company incorporated under the Companies
Act and having a place of business at
Kirkstane House 139 St Vincent Street
Glasgow G2 5JF, DEFENDERS.
(SC 858/08)
Edinburgh June 2009.

The Sheriff having resumed consideration of the cause Finds the following facts admitted or proved:-

[1]. The pursuer is aged 48 years. He resides at 20a Binny Park Ecclesmachan. He is an Information Technology specialist.

[2]. The defenders are an insurance company having a place of business at Kirkstane House 139 St Vincent Street Glasgow. Miss Wanda Milne is covered by a policy of insurance with the defenders to drive a motor vehicle registration number J30 BEN.

[3]. On 27th July 2007 the pursuer was driving his motor vehicle registration number M550 APU in Queensferry Road Edinburgh. He was stationary at traffic lights. The defenders' insured, Miss Wanda Milne, driving motor vehicle registration number J30 BEN collided with the rear of the pursuer's vehicle.

[4]. The collision was the fault of the defenders' insured.

[5]. As a result of the collision the pursuer sustained a flexion extension movement resulting in some tearing of the muscles in his left upper back and soreness in his left shoulder.

[6]. The pursuer attended his General Practitioner and was prescribed ibuprofen. He was referred to a physiotherapist. Between 27th August and 13th December 2007 the pursuer received about 20 sessions of manipulation, massage, acupuncture and laser treatment. His sleep was disturbed. Initially he suffered constant pain. His upper body movement was restricted.

[7]. The pursuer did not take time off work. He reduced the amount of driving he

would normally undertake. He was unable to resume his hobbies of yoga, woodworking and fishing until after he had completed physiotherapy.

[8]. He recovered within 12 months from the date of the accident although he continues to suffer some discomfort following periods of activity or inactivity.

THEREFORE grants decree against the defenders for payment to the pursuer in the sum of (First) Four Thousand Pounds (£4,000) Sterling with interest at the rate of 4 per cent per annum from 27th July 2007 until 26th July 2008 and at the rate of 8 per cent per annum from 27th July 2008 until payment; (Second) Forty Pounds (£40) Sterling with interest at the rate of 8 per cent per annum from the date of decree until payment and (Third) Seventy Five Pounds (£75) Sterling with interest at the rate of 8 per cent per annum from the date of decree until payment; Finds the defenders liable to the pursuer in expenses and fixes a hearing for the assessment of expenses to take place within the Sheriff Court House 27 Chambers Street Edinburgh on

NOTE:

[1]. The pursuer claims damages for the loss injury and damage caused by the defenders' insured in the accident on 27th July 2007. Liability was admitted. Prior to proof a joint minute of admissions was lodged whereby it was agreed that the medical report by Dr WA Campbell dated 18th (sic) March 2008 was to be treated as the medical evidence in the case, that the pursuer had recovered within the time period stated in Mr (sic) Campbell's prognosis, namely within 12 months from the date of the accident, that the pursuer had suffered loss of use in the sum of £40 inclusive of interest and inconvenience in the sum of £75 inclusive of interest. Only the amount of solatium was in dispute.

[2]. On 27th July 2007 the pursuer was driving his motor vehicle registration number M550 APU in Queensferry Road Edinburgh. While he was stationary at traffic lights the defenders' insured driving motor vehicle registration number J30 BEN collided with the rear of his vehicle.

[3]. The pursuer gave unchallenged evidence about the consequences of the accident. Initially he suffered soreness to his left shoulder and left upper back. He had limited movement and constant pain. He attended his GP and was prescribed ibuprofen. On a scale of 1 to 10 the pain was about 8. He did not take time off work but restricted the amount of driving and worked more from home. His area of responsibility was the whole of Scotland. He was referred to a physiotherapist and underwent about 20 sessions of manipulation, massage, acupuncture and laser treatment. At the conclusion of the physiotherapy sessions the pain was about 5 on the scale of 1 to 10. It was agreed with the physiotherapist that further sessions would be of minimal value. His sleep was disturbed. He had difficulty driving because of his restricted movement. He was unable to pursue his hobbies of yoga, fishing and woodworking until he had completed the physiotherapy sessions. He had resumed his yoga on a restricted basis and could carry out his woodworking with assistance in lifting. By March 2008 the pain had reduced to about 2 on a scale of 1 to 10. He continued to suffer pain in his back particularly after a period of inactivity or activity. He was able to tolerate the level of discomfort.

[4]. Dr WA Campbell, General Practitioner, examined the pursuer on 18th March 2008. His report dated 28th March 2008, which was agreed, states that the pursuer suffered a flexion extension movement resulting in some tearing of the muscles in his upper back. On examination there was some tenderness over the muscle below the left scapula but movements were otherwise full. Prolonged physiotherapy was considered to have been successful and the pursuer was largely symptom free. No long term sequelae were expected.

[5]. It was agreed that the pursuer had recovered within 12 months from the date of the accident although on the basis of the pursuer's evidence the pursuer's agent submitted that recovery was to a nuisance level rather than full recovery.

[6]. The pursuer's agent submitted that a reasonable value for solatium was the sum of £4,000, that interest should be applied at the rate of 4% from the date of the accident for a period of 12 months and at the rate of 8% thereafter. In support of her submission I was referred to the following:-
   1)   Urqhuart-v-Coakley Bus Co Ltd 2000 GWD 27-1047,
   2)   McGuire-v-Nicholson 6 November 2002 Unreported (Sheriff Jessop Stonehaven),
   3)   MacDonald-v-Bruce 8 August 2008 Unreported (Sheriff Evans Cupar),
   4)   Moir-v-Wilson 1 July 2002 Unreported (Sheriff Mackay Kilmarnock),
   5)   Spencer-v-Baron 4 February 2008 Unreported (Sheriff Morrison Edinburgh),
   6)   MacQuarrie-v-McKinstray 2007 SLT(Sh Ct) 120,
   7)   The Judicial Studies Board Guidelines for the assessment of General Damages in Personal Injury Cases (9th Edition).

[7]. With regard to the Judicial Studies Board Guidelines the pursuer's agent submitted that the pursuer's injuries fell between a moderate neck injury and a minor back injury resulting in a range of awards between £2,750 and £5,000. She submitted that the pursuer had made a fairly protracted recovery with residual nuisance level discomfort.

[8]. The defenders' agent submitted that a proper value for solatium was in the range between £1800 and £2250. She agreed that interest should be applied at the rate of 4% for a period of 12 months and thereafter at the rate of 8%. In support of her submission I was referred to the following decisions:-
   1)   Hall-v-Cockburn 16 February 2009 Unreported (Sheriff Hammond Ayr),
   2)   Sharp-v-Watt 19 March 2008 Unreported (Sheriff Muirhead Linlithgow),
   3)   Fairley-v-Thomson 2 September 2004 Unreported (Sheriff Allan Edinburgh),
   4)   Valentine-v-McGinty 20 May 2008 Unreported (Sheriff Kinloch Linlithgow),
   5)   Traynor-v-Kidd 1 August 2008 Unreported (Sheriff Paterson Dundee).

[9].   With regard to the Judicial Studies Board Guidelines the defenders's agent submitted that the Guidelines had not been "in force" at the time of many of the decisions referred to on behalf of the pursuer. Subsequently she agreed that they were no more than guidelines.

Discussion:

[10]. The only issue in dispute at proof was the amount of damages for the pain and suffering of the pursuer as a result of the injuries sustained by him in the accident on 27th July 2007. Liability was admitted. Dr Campbell's report dated 28th March 2008 was agreed as the medical evidence in the case. Further the amount of damages for loss of use and inconvenience was agreed.

[11]. The pursuer gave evidence in a straightforward manner. There was no attempt to embellish or exaggerate. It was his evidence that he continued to suffer discomfort after periods of activity or inactivity at a pain level of 2 in a scale of 1 to 10. That implied that there had been little, if any, improvement in his condition since the examination by Dr Campbell in March 2008. However it was agreed that "the pursuer recovered within the time period stated in Mr (sic) Campbell's prognosis; namely within 12 months from the date of the accident". There is no time period for full recovery contained within Dr Campbell's report. All that is stated is that he does not "expect any long term sequelae from this accident". A letter dated 16th May 2008 from Dr Campbell in which it is stated that "If this improvement was maintained I would expect that within 4 months of my examination ie by the middle of July he should be free of symptoms" is lodged in process. However no witness spoke to its terms nor were they the subject of agreement. Nonetheless it is clear from the terms of the joint minute of admissions that parties have agreed that the pursuer recovered within 12 months from the date of the accident. Perhaps recognising the inconsistency between the terms of the joint minute and the pursuer's evidence the pursuer's agent sought to suggest that "recovery" may not mean full recovery but recovery to a nuisance level. If "recovery" was intended to be qualified parties had the opportunity to express any such qualification and have not done so. It is also significant that parties were also agreed in the treatment of interest whereby interest at 4% is to be applied for a period of 12 months and thereafter interest at 8% is to be applied. In my opinion it is clear that parties are agreed that solatium is all in the past and all in the period of 12 months from the date of the accident. Accordingly, notwithstanding the evidence of the pursuer, in terms of parties' agreement damages require to be assessed on the basis that the pursuer recovered from the consequences of the accident within a period of 12 months.

[12]. It has been said on many occasions that the purpose of an award of damages is to compensate the pursuer in so far as money can for the loss suffered as a result of the accident. Each case requires to be considered on its own facts and circumstances. Each individual's reaction to and the consequences of an accident will inevitably differ. While there may be some similarities between cases there are also likely to be as many differences, such as the ages of the pursuers, the nature of their occupations, their levels of fitness and range of normal activities, their resistance to pain and their attitude towards medical intervention. While no two cases are identical "justice requires that there be

consistency between awards" as Lord Donaldson said in his foreword to the first edition of the Judicial Studies Board Guidelines in 1992.

[13]. The JSB of England and Wales produced the Guidelines with a view to assisting Judges in the difficult task of assessing the amount of damages in an action where personal injury had been suffered. They were not intended to be a ready reckoner or to fetter judgement in the particular case. The Guidelines are, as Lord Justice Waller said in his foreword to the ninth edition, a "framework for the assessment of damages in personal injury cases". Each edition of the Guidelines has taken account of the impact of inflation and decisions reached subsequent to the issue of the previous edition.

[14]. Despite the somewhat inexplicable initial suggestion by the defenders' agent that the Guidelines were not "in force" at the time of some of the authorities referred to by the pursuer's agent it was accepted that the Guidelines are no more than guidelines. It is clear from the authorities produced that the Guidelines are not always referred to and where they are referred to they may or may not be influential.

[15]. The framework of the Guidelines is to identify different types of injury and then to categorise them in terms of severity providing a range of awards for each category of each type of injury. As can be seen in this case, some injuries do not fit easily into one category. The pursuer's agent's submission that the injury suffered by the pursuer fell somewhere between a moderate neck injury and a minor back injury was not challenged by the defenders' agent. According to the pursuer's agent that would produce a range of awards between £2750 and £5000, although in the Guidelines that range appears to apply to minor soft tissue and whiplash injuries where the symptoms are moderate and a full recovery takes place within about two years.

[16]. Both agents referred to a number of largely unreported decisions which were considered by them to demonstrate the level of awards made in similar cases. It was by no means a comprehensive review of decisions, which may be wholly impractical particularly where unreported decisions are also relied upon, nor was it a review of decisions made since the ninth edition of the Guidelines was produced. It is unsurprising that the decisions referred to by the pursuer resulted in higher awards than those referred to by the defenders. The awards range from £1910 to £3780 allowing for inflation.

[17]. It is neither necessary nor helpful to analyse each decision to which I was referred. None of the cases appears to me to be directly comparable to the circumstances in this case. While in all the cases the pursuer is said to have suffered a whiplash type injury the nature of that injury and the pursuer's reaction to it differs, as is to be expected. The injury sustained by the pursuer in this case appears to me to be more severe than that described in the decisions to which I was referred. The defenders' agent appeared to me to attempt to minimise the extent of the injury sustained. Dr Campbell reported that the flexion extension movement caused by the accident resulted in some tearing of the muscles in the pursuer's upper back. That appears to me to be significant and describes a more severe injury than those described as a soft tissue injury in the

decisions referred to which may be more in the nature of bruising of the tissue rather than a tearing of muscles. Dr Campbell also reports that the pursuer underwent an "extensive" course of physiotherapy from which I infer that the physiotherapy undertaken was more than might be considered the norm. The pursuer's evidence was that after about 20 sessions a point was reached whereby further sessions would be unlikely to bring about any further recovery. He was prescribed ibuprofen and his sleep was disturbed. As with many individuals, particularly those who are self-employed, the pursuer continued to work notwithstanding the pain and limitations caused by the injury but he was unable to do the amount of driving he would normally do. His hobbies were restricted completely for about 5 months and thereafter he was able to reintroduce them although still to a lesser extent than prior to the accident and in the case of his woodworking with assistance. Some 8 months after the accident the level of pain suffered had reduced from 8 out of 10 to 2/3 out of 10. The pursuer had largely recovered in about one year from the accident. It may be that the pursuer's hobbies in particular yoga may have contributed to the speed with which he did recover.

[18]. In all the circumstances of this case and having regard to the decisions to which I was referred and the Judicial Studies Board Guidelines I consider that an appropriate award of solatium is £4000. As agreed between parties interest will run on that sum at 4 per cent from the date of the accident for one year and thereafter at the rate of 8 per cent.

[19]. It was also agreed that expenses would follow success. A hearing for the assessment of expenses will be fixed unless these are capable of being agreed.

IN THE FAMILY PROCEEDINGS COURT

Mr. H-D for the Applicant,
Miss P for the 1st Respondent,
Mr T 2nd Respondent,
Mr B 3rd Respondent.
Justices' Reasons.

1. The court is concerned with the one child, "A", approximately 8 months old. A is said to be not a particularly well child, although the extent of this is uncertain as A will be the subject of ongoing future medical testing. Mother is in her forties. Father is in his fifties. The parents are married. A is Mother's fifth child but the second child of both Father and Mother.

2. The Local Authority has applied for a care order and a placement order in respect of A. Care plans have been filed. The care and placement applications have been consolidated within these proceedings. The Local Authority seeks a suitable adoptive placement. This was envisaged from a very early stage.

3. These proceedings commenced on 1st May 2009, an interim care order was made on 5th May 2009 and has been renewed on appropriate dates through to this final hearing. A is placed with foster carers with whom A has remained during the course of these proceedings, in effect since birth. The parents have exercised contact throughout these proceedings. This is currently supervised contact, 5 days a week at 2 hours per session.

4. The parents opposed threshold, the care order and the placement order sought. They opposed the care plan. They presented as a couple (although separately represented) and wished A returned to them, indicating through submissions of their respective Counsel, in the nature of witness cross examination and in their written statements that they would cooperate with any support services in future. There was no alternative fall back carer put forward by the parents.

5. The Guardian supported the Local Authority's applications and the care plan.

6. The position of the parents changed by the second day of the final hearing. Both Mother and Father were to give evidence on the second day. On the second day Mother's Counsel stated that the parents did not wish to continue by giving evidence. They did not consent to the applications but did not oppose any more than to the extent they did at the end of the first day. The parents felt the giving of evidence would not take their case further. Father's Counsel indicated that the parents had reflected on the evidence heard on the previous day. It was said Father was unlikely to persuade the court to take a different view to the evidence of the Independent Social Worker. He, Father, knew the court will inevitably make a care order, then a placement order – then to an adoption order. He understood the consequences. He no longer opposed the applications.

7. The Threshold criteria the Local Authority wishes to establish is set out in the Amended Draft Threshold Criteria document dated the 27th November 2009. The Local Authority contends the threshold criteria under Section 31 Children Act are satisfied in that on the relevant date 1st May 2009 the child was likely to suffer significant harm and the likelihood of harm was attributable to the care likely to be

given to the child if the Order were not made not being what it would be reasonable to expect a parent to give to a child. This document was amended at the beginning of the second day of final hearing in that the Particulars at paragraph 1 were amended and paragraphs 6, 7 and 8 were abandoned by the Local Authority. The Particulars therefore finally relied on were:-

1. A's half siblings, H and R were made subject to full care orders made on 15th August 2005 with a care plan of long term foster care, The reasons for these proceedings was that H and R suffered significant harm by being sexually abused by different male associates of their mother, with their mother's knowledge as contained in the Agreed Threshold document dated 17th August 2005.

2. Within those Care Proceedings, a number of assessments were carried out, none of which were able to recommend the return of (Mother's) children to her care. These assessments were:

   1. A Psychological assessment completed by the Chartered Consultant (dated 9th May 2005), who concluded that (Mother) had allowed herself to be sexually exploited because she "did not have the ability to form judgements about the appropriateness of them." Also that she was "unlikely to learn appropriate parenting skills within the developmental timescales of her children."

   2. A Psychiatric report, completed by the Consultant Psychiatrist (dated 6th July 2005). Who commented that "I am afraid that I would not have confidence that (Mother) can parent any of her children well enough to protect them from abuse or other kinds of harm."

   3. On 19th June 2006 a Care Order was granted in respect of S who was born during the proceedings relating to H and R. Within S's Care Proceedings, a Psychological report was compiled by a Chartered Psychologist (dated 7th October 2005). The Psychologist concluded that "I do not believe that (Mother) could protect herself or any child in her care from further sexual abuse."

   4. K was born on 21st June 2007 and proceedings were commenced immediately and came to a conclusion on 24th January 2008 with the making of care and placement orders. There was a contested hearing and the Justices made findings (which were contained within the papers filed with the court).
   We do not intend to repeat all those findings, as they run to several pages but findings in relation to both parents were, hostility to professionals and lack of insight to Local Authority concerns. In relation to Father findings were violence and threats of violence to his family, violence to mother, violence to neighbours and their children.

   5. The Court heard from the chartered psychologist (referred to in 2.1 above) and the following remarks from him were quoted in their judgement
      (a) "The only way that the risks could be managed was through a package of waking hours (support) 7 days a week until K attained the age of 18 or at least into his teens."
      (b) That if the allegations of violence against the father were proved "(Father) and (Mother) are probably too risky to be trusted with the care of their child.

(c) The parents would always be playing catch up with the development of their child.

8. As is seen from the Threshold sought to be satisfied, there have been previous care proceedings. They have been in relation to Mother's four elder children:- R, H, S and K. Care orders were made in respect of R and H on 15th August 2005. Both are long term fostered. S was born during those care proceedings. A care order was made for S on 19th June 2006. He was adopted on 11th May 2009. Only K is the child of Father in these proceedings.

9. K who is Father and Mother's first child together was placed into foster care the day following his birth. On 24th January 2009 a care order and placement order was made for K. On 11th May 2009 K was adopted.

10. Mother has not had an easy life. Psychological assessments in previous proceedings indicate that her abilities place her in the category of learning disabled. She suffers from V W disease preventing blood clotting properly. She suffers from asthma and, it is said, carries excess weight. She was known to the Local Authority since before R's birth. In early 2003 when living in another part of the country with R and H it came to light those children were being sexually abused by male associates of Mother. R and H told Mother of the abuse but she continued to associate with the males, to bring the children into contact with them and permit them to care for the children without her being present. Mother breached a written agreement with the Local Authority when the children were returned to her care. In the fullness of time, as said, care orders were made for R and H on 15th August 2005. In those proceedings, the psychological assessment of Mother concluded "(Mother) has very limited intellectual ability. She has been unable to perceive the risks to her children – in spite of her own experiences of similar abuse throughout a substantial part of her life". Further, "(Mother) is unlikely to learn appropriate parenting skills within the development timescales of her children. She is unlikely to be able to provide safe parenting against offenders who are commonly skilled in identifying vulnerable people".

11. In those 2005 proceedings a parenting assessment of Mother was undertaken by a Consultant Psychiatrist and, despite Mother undergoing protection work with the Lucy Faithful Foundation, he agreed with the said psychological assessment of Mother in relation to Mother's ability to protect and her potential for change. He concluded she was not able to protect R and H from emotional and physical abuse. "Sexual abuse is only one of several risks to children in (Mother's) care. She has difficulty with coping with life's difficulties in general and her description of home circumstances when H and R were living with her indicate a very impoverished and neglectful style of parenting. Her account of (R and H's father's) behaviour towards the children indicates she was not able to protect them from physical and emotional abuse. She has not been able to protect herself and would not be able to protect a child in her care."

12. In S's proceedings the psychological assessment concluded "Therefore I do not believe that (Mother) could protect herself or any child in her care from further sexual abuse".

13. Father has come to fatherhood late in life. He has had no substantive experience of parenting, both his children including A being removed at birth. He is said to have had a comparatively solitary lifestyle, little socialisation and found his partner late in life. He has his own health problems. He has impaired mobility. The psychological assessment in K's proceedings indicates Father only has a slightly higher rating than Mother in the category of learning disabled.

14. In the care proceedings involving the parents' first child together, K, the Psychologist's addendum assessment of 8th November 2007 expressed the view "Indications are that (Father and Mother) may both have learning disability to a greater or lesser extent has rendered it unrealistic to employ detailed psychometric assessment often used to investigate personality patterns and potential disorders". The effect of the learning disability for Mother has been to leave Mother vulnerable to exploitation, difficult to acquire new skills and Mother would find it hard to keep up with K's developmental needs. Even if Mother and Father acknowledge support is needed and is asked for the help the parents would need is considerable. The assessment stated "They are likely to require support throughout periods when they are interacting with the child. This will amount to waking hours, seven days a week……." Given Father's impulsive aggression, if this were established the parents were probably "too risky to be trusted with the care of a child".

15. The independent parenting assessment concluded each parent was not equipped to care for a child alone. If domestic violence was a feature of the relationship, the presence of Father would not overcome Mother's shortcomings. Even if domestic violence could be disproved the author was "not particularly confident that the couple have sufficient personal resources to meet (K's) needs effectively…" There were too many risk factors.

16. In K's proceedings the Justices did make findings as in the Reasons of the 24th January 2008 which we do not repeat here in full, but emphasise, from page 4 at Paragraph Ba-Bj "…we have already proven that (Father) has a propensity to commit violence…."

17. Given the family history and the way the Threshold document as amended was formulated we considered it important to set out the duration of the problems and refer to attempts to previously support the family. The Local Authority's concerns are a direct consequence of the past events which led to A's removal at birth.

18. The precipitating event for these proceedings was A's birth.

19. The parental conduct since birth has, it is accepted, been generally good. Both parents have engaged well, have been regular in attending their supervised contact for 2 hours 5 days a week and the quality of contact has been good.

20. At this hearing we heard oral evidence from the Social Worker, and the Independent Social Worker, whose instructions were led by the parents but who was called by the Guardian. The evidence of these witnesses was challenged in cross examination by the parents. The Guardian was not called to give evidence. Her written reports were considered. Her evidence was not challenged by the parents. The parents did not give oral evidence in the circumstances we have described above at paragraph 6 and repeat here. Effectively, the parents cannot consent but do not actively oppose.

21. We read the bundles submitted including the reports therein. We have had handed up the original reports from the previous proceedings. The previous proceedings Reasons/Court findings were not actively challenged save for the said disagreements with these referred to in the written statements of the parents.

22. We were not referred to any statutes nor case law.

23. The issues to be determined at this final hearing are whether the Threshold criteria are satisfied pursuant to Section 31 Children Act 1989, is there sufficient information available to this court to enable an accurate conclusion to be reached as to the capacity of A's parents to provide A with good enough parenting, and their ability to demonstrate they have taken on board the professionals concerns and that

they have the capacity to change or have achieved change in order to address those concerns so that A can be provided with good enough parenting.

24. As to the Local Authority evidence, it relied on previous, although relatively recent, psychological and psychiatric assessments in the previous care proceedings some of the conclusions of which are recorded above. While the parents have not accepted these assessments and dispute the findings in their statements they remain unchallenged by the parents who gave no oral evidence. The parents have not produced any contrary expert evidence to refute these findings. The Findings/Reasons of the court in previous proceedings, again, were unchallenged. These, however, are matters of record.

25. The evidence of the Social Worker was that she has not seen any significant change in the parents to alleviate the Local Authority's concerns. She concentrated on four criteria. Firstly, the parents inability to work with agencies. Mother did not disclose her pregnancy until late. It was, however, accepted that all ante natal appointments were kept once "booked in". Secondly, historical concerns of sexual abuse. The parents showed no understanding of previous concerns. There was no further insight by the parents. There was insufficient change there. Thirdly, the parents' relationship. Positively for the parents, the fact they were still together was a good thing. There had been no recent complaints to the Police. This was qualified in that she only saw the parents on limited occasions. She was concerned the relationship was sustainable enough for a longer period. She recalled the correspondence of Mother writing to Social Services saying she was scared of Father who threatened to kill her. There was no input from outside agencies sought by the parents. Fourthly, the support from the Local Authority or other agencies. When the parents were asked by the Social Worker about support, they would respond that they would accept any support given. The parents had not approached the Social Worker at all to request any support. There was no insight by the parents as to what support would be needed if A went home. Nothing was said of A's potential special needs. The parents had shown antagonist views of the Local Authority and towards the foster carers. The Social Worker confirmed A's health was not good and that most things would be "found out". She agreed with the Independent Social Worker's report in that A needs a high level of care. In short, her evidence was that the parents had not changed sufficiently so that it was safe to return A to the care of the parents.

26. Cross examination of the Social Worker did reveal that the decision of the Local Authority to plan for long term adoption was made at an early stage based on earlier assessments from previous proceedings. Further, it was revealed that the Social Worker did not discuss the final care plan with the parents. The contents were conveyed through solicitors. Her reason for this was that she was told that the parents had made a complaint about her, the exact nature of which was unknown to her. As to this last point, we consider the parents could have been treated more sympathetically in a face to face discussion of the final care plan, but our findings do not, in any way, turn on this point.

27. The only other oral evidence we heard was from the Independent Social Worker. She was instructed by the parents to carry out an assessment of the parenting abilities of the parents. A report and supplemental report were prepared, the outcome of which was that she supported the position of the Local Authority and not the parents, in that the final recommendation was that A is not placed into the care of the parents. For that reason the parents did not call her as a witness. She was called by the Guardian.

28. The Independent Social Worker stated the parents are in a loving, stable relationship. It is a positive relationship and they are interdependent on each other. Their current home is clean and physically suitable for a child. There is, however, no internal challenge to the relationship. Their commitment to A in contact and the handling of the baby was good. There is no doubt they love A. The main level of concern was their intellectual abilities. Into her overall considerations were the special needs of A. Child A is not a robust child. Mother, in her judgement, did not appreciate the risks of sexual abuse, even after having the support of the specialist Lucy Faithful Foundation Mother had to take some responsibility. Father shows some awareness as to the risks of a child in the parents' care generally but it is documented that when challenged he can become very angry. There were indicators of Father's potential for confrontation. If he reverts to previous behaviour it is an unsuitable environment for A. He has not had anger management, he did not feel the need for it. Father is assessed at an intellectual ability not hugely above that of a child. The parents were untruthful when they said they had given up smoking but had not. A is sensitive to smoking. Mother has her own health difficulties and does not look after her own health needs. There was no direct evidence of A's special needs, a diagnosis is awaited for A's dismorphic features. He sometimes stops breathing. Something is "not quite right" with him. He needs extra handling and sensitivity.

29. The reports of the experts in the previous proceedings had been considered by the Independent Social Worker in carrying out her thorough assessment. She states there was no evidence of a change from the findings in those reports. The psychologist's addendum report in 2007 said the parents would need twenty four seven support. The parents have been doing their best, but they have limitations and they are struggling. She has concerns that the parents' relationship can survive A being returned and considers that neither can parent safely individually. There are difficulties, the IQs of the parents cannot be increased, they have health difficulties, there are risks. Not enough evidence has been provided to show that the parents have changed sufficiently to provide an appropriate level of care to A.

30. The Independent Social Worker considered a residential assessment of these parents with limitations. It was a difficult decision not to recommend one. It was not disputed the parents could provide a basic level of care but a baby growing up in a non residential assessment setting is a far cry from a residential assessment. The parents were already working to the best of their abilities and any future change would be small. The concerns remained unresolved. The parents cannot grasp risk. She was adamant a residential assessment would not help.

31. The parents did not give oral evidence, so their evidence could be challenged. We did consider the parents' written statements. In short, the parents accepted they had learning disabilities, did not agree with the views of previous expert reports, did not agree with findings of the court in previous proceedings and state that they have changed sufficiently and would accept all appropriate assistance so that A could be safely returned to their care.

32. The Guardian's evidence was contained in her Reports in respect of the care order application and a further Report in respect of the placement order application. She supports the Local Authority applications. All assessment reports of Mother and Father are not positive and indicate they are not in a position to safely care for A. Granting the orders would ensure A's long term welfare. The care plan is supported. The evidence of the Guardian was not contested by the parents.

33. Our conclusions in findings of fact are hereafter.

34. We find that as fact the parents cannot contest the agreed Threshold document dated 17th August 2005 as to the knowledge of mother of sexual abuse in relation to H and R as contained therein. We accept the assessments and findings of the Consultant Psychologist dated 9th May 2005 and the Consultant Psychiatrist dated 6th July 2005 that Mother could not protect her children nor likely to learn appropriate parenting skills as said therein. Likewise, we find the findings of the report of the Chartered Psychologist of 19th June 2006 in relation to S's proceedings cannot be contested. The findings of the court in K's proceedings cannot be contested, including the quoted remarks of the Chartered Psychologist.

35. We found the evidence of the Social Worker, the Independent Social Worker and the Guardian persuasive and accept this evidence. We find A has needs more than that of a normal child, even though A's needs have not been fully determined by medical practitioners. His parents in written evidence accept this. We accept that Mother and Father love A. They have tried their best within their limited intellectual abilities. Both Mother and Father have health problems. They both have learning difficulties. We accept they are in a stable relationship but we have concerns as to their relationship if A were returned to them. In that case there would be a substantial risk to the relationship. Neither parent could parent on their own. There is a long history of Mother failing to protect. Mother does not appreciate the risks of sexual abuse, even though she has received assistance to try to appreciate this. The parents have shown no real understanding of previous concerns of the Local Authority. The parents do not have the ability to protect A. Father has had findings of a propensity of violence and threats made in the past. There is a very real concern that he could revert to previous behaviour. The parents have been slow to engage the support of agencies. The parents have not actively been able to request support unless initiated by the Local Authority. There has been no real insight into the assistance needed should A with all A's needs return to parents' care.

36. We accept all the assessment evidence relied on by the Local Authority from previous proceedings. The concerns of the Local Authority are unresolved. We accept the parents cannot grasp risk. Any future change would not be enough. The parents have not changed sufficiently to provide an appropriate level of care. They could not jointly or individually care for A.

37. As to Threshold we are satisfied pursuant to Section 31 of the Children Act 1989 that on the relevant date 1st May 2009 the child was likely to suffer significant harm and the likelihood of harm was attributable to the care likely to be given to the child if the order were not made not being what it would be reasonable to expect a parent to give to the child. As indicated above Paragraphs 1, 2, 3, 4 and 5 are agreed documents, assessments and findings already made and accepted. The parents in our judgement cannot dispute this. There was no contrary evidence provided by the parents.

38. The Threshold criteria having been satisfied, we turn to whether there is sufficient information available to this court to enable an accurate conclusion to be reached as to the capacity of A's parents to provide him with good enough parenting. We mention this since the Independent Social Worker was questioned as to whether a residential assessment in particular should have occurred or could occur in future. We accept the evidence of the Independent Social Worker in this respect that it was not appropriate for the reasons she gave in evidence set out above. We find the court has sufficient information available to it to enable an accurate conclusion. We would also add, that save for the cross examination points raised by Father's

Counsel as to why a residential assessment was not undertaken – no submissions were made as to this point nor did the parents give oral evidence in relation to it.

39. We now consider if there is a need for an Order and if so, which Order. We have the child's welfare as our paramount concern. In doing so we address the Welfare Checklist in Section 1(3) of the Children Act 1989.This has been fully addressed by the Guardian and we agree and adopt her assessment of 14th December 2009 as our own. In particular we mention: - at Paragraph 7(d) the health issues of A and possible potential illness and/or disability; at Paragraph 7 (e) as to harm, the history of Mother and Father suggesting they are not in a position to safely parent A; at Paragraph 7(f) as to the capability of the parents, the expert assessments referred to therein and the assessment of the Independent Social Worker recommends A is not placed in the care of his parents as set out.

40. The significance of applying the Welfare Checklist is that Mother and Father are not in a position to care safely for their son. There are no other family members offering care to A.

41. We have considered the full range of powers/orders available to the court.

42. We have considered the least interventionist "No Order" principle and whether it would be applicable. In this case, however, in accordance with the findings in conclusion we have reached it is clearly appropriate for an order to be made. It would not be safe for A to return home

43. The order we are making in our judgement has to be a Care Order. The child is not being placed within the family. No other type of order is appropriate.

44. We approve the final care plan of the Local Authority recommending A be placed for adoption including the pattern of contact as set out therein.

45. We make a Care Order to The Local Authority.

46. We now turn to the application for a Placement Order pursuant to section 22 Adoption and Children Act 2002. This order would authorise the Local Authority to place the child for adoption with any prospective adopters who may be chosen by the Authority. Neither parent has given consent. The position of the parents to the Placement Order application is set out above. The court can only dispense with the parents' consent if the welfare of A requires the consent to be dispensed with. In reaching our decision we have had regard to the findings set out earlier in our judgement in respect of the evidence.

47. We have carefully considered the criteria in Section 1 of the Adoption and Children Act 2002. We remind ourselves that the paramount consideration of this court must be the child's welfare throughout A's life and that in general any delay in coming to a decision is likely to be prejudicial to A's welfare. We have again considered a full range of powers under the 2002 Act and under the Children Act 1989 and we must not make any order unless it would be better for the child than not doing so. We have addressed the Welfare Checklist under the said 2002 Act. The Guardian addressed this in her report dated 13th January 2010 in the placement application. We agree her findings in this regard and adopt them in our Reasons. The Guardian states A will require a planned move to prospective adopters should be completed as soon as possible. A Placement Order will provide permanence and stability for A.

48. We find an adoptive placement is the only placement that would provide the stability and security that meets A's needs throughout A's childhood.

49. Neither parent gave any oral evidence opposing the placement application.

50. Given our findings we are satisfied the child's welfare requires us to dispense with the consent of the parents which we do. We have found the child cannot safely be

returned to either of A's parents, no other family member can care for A and therefore at A's age the only appropriate placement is an adoptive placement.

51. Accordingly we make a Placement Order in respect of A and in doing so approve the contact arrangements.

52. We were not referred specifically to any Human Rights issues. In making the orders in this case the court has considered the rights of the parties and the child, in particular the right to a fair hearing and the right of any individual to enjoy family life. All Respondents have been legally represented and we are satisfied they have had a fair and proper hearing. The decisions we make are proportionate. The rights of the child to ensure he is protected outweighs the rights of the parents. The child's welfare is the paramount consideration.

53. The decisions we have made are in the best interests of A and these will be difficult for the parents who may not agree with the decisions. We would state that it has always been accepted in these proceedings that Mother and Father love A and have tried as hard as their abilities and learning difficulties allowed in their attempts to improve.

Lay Bench.
Legal Advisor Mr M.