UNIVERSITY OF MURCIA

International Journal of English Studies

IJES

www.um.es/ijes

Words for the Transitional Bubble: A Lexical Analysis of Two Economic Crises

Mª ÁNGELES ORTS LLOPIS
CAMINO REA RIZZO
Universidad de Murcia

Received: 10 June 2010 / Accepted: 3 December 2010

ABSTRACT

This study aims at the analysis of the lexicon in English of the two professional areas, telecommunications and finance, affected by the crises of the recent years: the 90s dot-com bubble and the present-day Credit Crunch. Both crises share a common context of wealth and cultural complexity, being the root for the coinage of innovative specialised terms and collocations. Our study is specifically aimed at unveiling the lexical coverage of both crises, in terms of technolects and their context, evolving in several phases. First, two corpora of specialised, semi-specialised and general texts from the domains' digital periodicals will be characterized according to lexical relevance and terminological volume, to see the extent in which they are lexically connected or diverge when experiencing a critical situation like a crisis. Finally, clarifying how far these two disciplines have related during the last critical years will hopefully provide some clues for the lexical ethnography of two institutionalised ways of thinking.

KEYWORDS:

Lexical analysis, dot-com bubble, credit crunch, crises, journalistic genre, digitized news.

RESUMEN

Este estudio se dirige hacia el análisis del léxico en ingles en dos áreas profesionales, la tecnología y las finanzas, muy afectadas por las crisis de estos últimos años: la de la burbuja tecnológica y la llamada crisis del crédito, en la que nos encontramos inmersos. Ambas debacles comparten un contexto común de riqueza y complejidad cultural, y han sido fuente inagotable de nuevas acuñaciones lingüísticas en forma de unidades y combinaciones léxicas muy sofisticadas. Nuestro estudio se dirige, especialmente, a estudiar la cobertura léxica de dichas crisis, con particular atención a los tecnolectos, y la manera en que estos evolucionan entre una y otra. Una primera fase del estudio reúne dos corpus de textos especializados, mixtos y divulgativos de procedencia digital y periodística, con el fin de dilucidar su relevancia léxica y su volumen terminológico y comprobando la red de conexiones y divergencias léxicas que se establecen en dos situaciones críticas diferentes. Finalmente, buscando relaciones léxicas entre estas dos crisis, esperamos suministrar claves que desvelen la etnografía lingüística de dos maneras de pensar del todo institucionalizadas.

PALABRAS CLAVE:

Análisis léxico, burbuja tecnológica, crisis del crédito, género periodístico, noticias en formato digital.

Address for correspondence: Mª Ángeles Orts Llopis. Departamento de Filología Inglesa, Campus de *La Merced* 30071, Murcia, Spain. Tel: 00 34 868883858; E-mail: mageorts@um.es. Camino Rea Rizzo. Departamento de Filología Inglesa, Campus de *La Merced* 30071, Murcia, Spain. Tel: 00 34 868883010; E-mail: camino.rea@um.es.

ISSN: 1578-7044

1. INTRODUCTION

The present work aims at the lexical study of two corpora reuniting a restricted set of genres deployed to describe the inception and development of the two most recent economic crises, i.e., the so-called dot-com bubble, and the present Credit Crunch. The onset of the millennium has successively seen the progress of these two dire economic catastrophes that have, paradoxically, been accompanied by an unprecedented era of technological expansion and financial sophistication. The first in time, the one labelled as the dot-com bubble, also known as the dot-com-crisis, refers to a four-year period (1997-2001) during which the Internet and technological sectors boomed and subsequently busted, due to the so-called *Network effect*. During the course of the debacle, companies in the hi-tech and telecommunications sector first became overvalued, moving swiftly and with little caution, by operating at a sustained net loss to build market share, creating unrealistic expectations in their investors. The ensuing bursting of the bubble brought an inevitable batch of sell orders, the collapse of the NASDAQ index and the splitting or destruction of many .com and e- companies, burdened with unredeemable debts and share devaluation.

The collapse of a global housing bubble and the lack of regulation in financial markets, together with predatory, fraudulent lending practices and an era of unparalleled consumption, have damaged global economies irretrievably, driving the world to a deteriorated state of exhaustion and debt.

Nevertheless, these two crises –with a chiefly technological and financial origin, respectively– which share a common context of wealth and cultural complexity, have also been the root for the coinage of an innovative and intricate plethora of specialised terminology. Such terminology has had as its source the English language as the lingua franca of communication, especially in science, technology and economics, in the recent decades. Indeed, as legal, economic and –to some extent– cultural barriers have been overcome by globalizing efforts, the world has witnessed the increasing influence of English as the dominant tool of interaction and common discourse for professionals in business settings and for business purposes (Palmer-Silveira *et al.*, 2008). The intelligentsia and think-tanks of business and science have, therefore, been educated in English, and have subsequently acquired Anglo-Saxon versions of the state-of-the-art terminology to analyse and express the most salient phenomena in their respective areas (Orts Llopis, 2005; Orts Llopis, 2007; Orts & Almela, 2009). As far as English for Business Purposes is concerned, the little research there is, has been carried out in the area of teaching, not in that of linguistic description. In the Spanish context, Posteguillo and Palmer's study is worth noting on Business Press genres (1996), as well as Alejo and McGinity's, in the field of incorporations into the Spanish language of economics (1997). Few, but substantial, studies have been developed about the nature and typology of economic terminology (Mateo, 2004; Nelson, 2000; Pickett, 1986), but no exhaustive examination has been accomplished of the words of the most important economic crises to date. Hence, the novelty of our present work, which aims at this unprecedented task as described above, namely the analysis of the lexical coverage in English of the two latest world economic crises: the dot-com bubble, and the Credit Crunch, as monumental periods of terminological coinage and verbal interaction. In doing so, we will exclusively concentrate upon two professional areas specially affected by these crises –technology and finance– with an aim to identify the most relevant terms deployed to verbalize the economic fiasco in each realm.

In accomplishing an evaluation of the most common specialised terms in the areas of technology and finance, our analysis will evolve in several phases. First, no effort has been spared in selecting a corpus to cover the lexis of the two economic phenomena in digital journals. In doing so, we deemed it necessary to involve the areas and time zones affected by the crises in our study, these being mainly the dot-com bubble, affecting essentially the technological and engineering sectors on the one hand, and the Credit Crunch, basically involving the financial, manufacturing and production industry sectors on the other hand. Thus, our samples –which we will respectively label hereafter as the dot-com corpus, or DCC and the credit crunch corpus, or CCC– mainly include specialised, semi-specialised and general texts in the form of specific genres: straight news reports from specialised publications and specialised sections with feature articles within general publications. Accordingly, the samples in the first subcorpus range from the years of the dot-com bubble, 1997 to 2000, and in the second is a compilation of texts from the inception of the present financial mayhem, 2007, to date, the year 2010. The dot-com corpus was extracted from the Telecommunication Engineering Corpus (Rea, 2008) whereas the credit crunch corpus has been compiled later to serve the purposes of the present study, being still in the process of construction, since it aims to give coverage to the crisis in its entirety. We have had to deal with the differences in sizes between the two corpora, one being the product of a closed crisis in the past, hence longer, the other of an economic fiasco as yet without foreseen end, thus incomplete. Even if there may be generic differences in audience and tenor between the samples (since they include genres aimed both at a specialised and a lay target market), we consider such differences insubstantial for the purposes of our analysis, as it primarily attempts to cover the recurrences in the usage at the lexical level of each of the subcorpora –namely technical terms in these specialised domains– and inspect to what extent these recurrences make the two subcorpora respectively connected or divergent, as far as their lexicon is concerned.

The intended evaluation and vocabulary detection is mainly aimed at proving that a useful method for the study and appraisal of the terminology is necessary for a non-English-speaking discourse community to understand the genres issued by the global think-tanks of each discipline, as voiced by the media, and, consequently, to acquire the discursive expertise necessary to provide information and feedback. In doing so, the subsidiary goal of this study will be to clarify how far technology and finance have related and developed during the last critical years, providing some clues for the lexical ethnography of two institutionalised ways of thinking, through the analysis of their respective genres.

2. METHOD

The samples in the corpora will be defined and identified in terms of terminological volume, by both executing Paul Nation's Range software (Nation & Heatley, 2002) and applying the quantitative parameters which condition term detection (Chung, 2003). Our study will try to detect the recurrences in the usage of technical terms in these specialised domains, and inspect to what extent they are lexically connected or diverge when experiencing a critical situation such as a financial or technological crisis.

Range allows sorting out the words in a text according to different vocabulary levels, on the basis of how often and how widely they occur in the language, that is, on frequency and range respectively. In most texts, around 80% or more of the running words come from the most frequent 2000 words of English (Nation, 2001). The word lists available for Range include (1) the most frequent 1000 words of English, (2) the second most frequent words, both from A General Service List of English Words (West, 1953), and (3) The Academic Word List (Coxhead, 1998, 2000). The lists contain the base forms of words and derived forms so that they register more than 12,900 types belonging to 2,086 word families.

When focusing on languages for specific purposes and definite, explicit genres, −such as economic news-items with specialised or general audiences− there exists a proportion of vocabulary that does not fall in the high frequency levels, but which is frequently and widely used within those specialized areas and/or genres. By running the corpora through Range, not only can we obtain a neat classification of the words coming from the first and second thousand word lists and the Academic Word List, but also a separate set of the words which do not come from any of these lists. The off-lists group gathers the subject-specific or technical vocabulary of the domain together with the lower-frequency units of the text.

As far as lexical coverage is concerned, 95% of the words are needed to be able to acquire an adequate global understanding of texts in specialised fields (Laufer, 1992; Nation, 1990). Of these, 2,000 high-frequency general service words are needed, together with 570 general academic word families and 1,000 or more technical words (Nation, 2001). However, written discourse is a heterogeneous phenomenon and may vary to some extent depending on a number of factors, such as register, level of abstraction, addressee, topic, genre, etc., so that the type of text has an enormous effect on the kinds of words used and therefore, on lexical coverage. For example, the Academic Word List made up an average of 10% coverage in a 3.5-million-word corpus of 28 different academic subject areas, whereas it managed to cover only 1.4% in fiction texts and a higher 4.5% of words in newspapers (Coxhead, 2000, 2006). In addition, subject-specific vocabulary may account for a substantial quantity of the words in a text, around 5% (Chung, 2003; Nation, 2001; Rea, 2008), although such figure fluctuates considering low-frequency words. Moreover, some specialized lexical units are context-dependant and take on particular meaning when used on different grounds. This is particularly the case in our corpus, as we will see below, and mainly in the keywords shared by both corpora.

Therefore, a corpus-comparison approach dealing with the statistical distribution of linguistic items becomes essential to sketch the lexical content characterizing each critical period. Hence, both subcorpora are contrasted to the general language corpus *Lacell*[1], in order to pinpoint those lexical units whose occurrence is statistically significant in comparison to their ordinary behaviour (as specific genre subcorpora deviate from the norm registered in the general corpus). The Keywords tool available in WordSmith programme (Scott, 1998) will reveal which words are drawn to the front out of stability periods and how the specific genre subcorpora deviate from the norm registered in the general corpus. In agreement with previous studies evidencing that such keywords tend to provide a clear account of the lexical content in texts (Nelson, 2000; Scott, 1997; Stubbs, 2001; Tribble, 2000), our analysis endeavours to find out to what extent the subjects in finance and technology domains coincide or differ particularly in times of economic and social crises, and whether the keywords they might share also behave in similar way.

3. RESULTS: STATISTICAL BEHAVIOUR OF THE LEXICAL CHOICE IN BOTH CORPORA

3.1. Basic statistical information: dot-com and credit crunch corpora

The first processing of both corpora by WordSmith yields basic statistical information on the composition of the samples with regard to tokens, types, type/token ratio and standardised type/token ratio (table 1). Corpus size is given by the number of tokens in every corpus, that is, by the total of running words. As many of the tokens are the repetition of the same sequence of characters, then the number of types or word-forms indicates the amount of different words in the corpus, including each form derived from a main lemma or headword. This set of types constitutes the vocabulary of the text.

Dot-com corpus	**Credit Crunch corpus**
Tokens: 405,357	Tokens: 265,008
Types: 49,422	Types: 27,177
Type/Token Ratio: 12.19	Type/Token Ratio: 10.26
Standardised Type/Token: 86.31	Standardised Type/Token: 88.53

Table 1. Basic statistical information.

As shown in table 1, the program counts 405,357 tokens/49,422 types in the dot-com corpus and 265,008 tokens/27,177 types in the credit crunch corpus. Although those figures reflect a considerable difference in size, their lexical diversity would compensate such unbalance, according to the ratios signalling the relationship existing between the total number of types and tokens. Type/token ratio is obtained from the division of the number of different forms by the number of running words and multiplied by 100. The higher the result is, the greater the lexical diversity of the sample. The credit crunch corpus obtains a ratio just 2 points lower than the dot-com corpus, which evidences a lower lexical burden in the text due to the repetition of the same forms. Likewise, the program computes the standardized type/token ratio every n words, being n=1,000. The ratio is calculated for the first 1,000 tokens, and then computed for the next 1,000, and successively until the end of the text, yielding the average of the obtained values. Again, there is just a 2-point difference in the outcome (DCC: 86.31 and CCC: 88.53). Therefore, their proportional lexical diversity is quite similar in spite of their raw size.

3.2. Keywords, ranges of specialty and ranges of frequency

The language samples are subjected to several tests which generate a wide range of quantitative data for every word. Among them, a first selection is displayed in table 2 for the dot-com corpus and in table 3 for the credit crunch corpus: keywords, frequency index in our samples, frequency index in the general corpus, specialty index (ratio and term) and keyness index.

The degree of relevance, or keyness, is given by the log likelihood test in WordSmith. This tool identifies keywords on a mechanical basis by comparing patterns of frequency. A keyword is defined as "*a word which occurs with unusual frequency in a given text*" (Scott, 1997: 237), that is to say, a word whose frequency is unusually high (positive keywords) or low (negative keywords) in comparison to a general norm. 2,311 keywords have a significantly higher frequency in the dot-com corpus, where the highest keyness value associated to a word is 8,991 (*network*) and the lowest one is 23.9 (*welcome*). The credit crunch corpus, in turn, gains 1,196 positive keywords whose indexes spread from 24 (*rated*) to 3,225 (*banks*). The set of statistical features of the samples defines the journalistic genre against the general language depending on the variation in the lexical choice, so that the meaning of lexical items is interpreted in discourse both by what they express and what they exclude. However, the current study focuses on the words that, statistically, are more probable to occur during the crises. Moreover, positive keywords usually provide a good account of the subject content: "*positive keywords give a good indication of the text's aboutness*" (Scott, 1998: 63).

WORD	Freq	Freq. Lacell	Ratio	Term	Keyness
NETWORK	1,719	1,686	52.86	SPC	8,991
IP	813	20	2,10	SPC	6,264
LINUX	701	16	2,27	SPC	5,410
SOFTWARE	908	1,412	33.34	NO	4,154
WIRELESS	569	171	172.51	SPC	3,722
DATA	1,005	2,787	18.69	NO	3,697
SERVER	580	362	83.06	SPC	3,361
NETWORKS	558	463	62.48	SPC	3,039
INTERNET	575	910	32.76	NO	2,615
APPLICATIONS	550	934	30.53	NO	2,443
WEB	520	791	34.08	NO	2,395
USERS	554	1,144	25.10	NO	2,295
STORAGE	502	803	32.41	NO	2,275
TECHNOLOGY	720	2,794	13.36	NO	2,256
ETHERNET	311	37	435.79	SPC	2,233
VPN	286	5	2,965	SPC	2,219
SERVERS	321	79	210	SPC	2,152
SYSTEMS	714	3,000	12.33	NO	2,145
ACCESS	680	2,696	13.07	NO	2,107
VENDORS	301	81	192.66	SPC	1,997
OPTICAL	323	164	102.11	SPC	1,947
DEVICES	379	476	41.28	NO	1,851
PRODUCTS	604	2,480	12.62	NO	1,837
MANAGEMENT	665	3,393	10.16	NO	1,787
COMPANIES	654	3,505	9.67	NO	1,705
CUSTOMERS	464	1,319	18.23	NO	1,688
VPNS	211	3	3,646	SPC	1,642
BASED	724	5,193	7.22	NO	1,546
SYSTEM	912	8,707	5.43	NO	1,538
PRODUCT	490	2,080	12.21	NO	1,463
SAYS	1,005	11,193	4.65	NO	1,461
BANDWIDTH	198	20	513.27	SPC	1,438
LAN	201	27	385.96	SPC	1,430
TELEWORK	181	2	4,692	SPC	1,414
XILINX	178	0	∞	inf/spc	1,412
SERVICES	700	5,742	6.32	NO	1,346
SECURITY	555	3,340	8.61	NO	1,342
MOBILE	303	526	29.86	NO	1,336
COMPANY	713	6,158	6.00	NO	1,313
DESIGN	546	3,313	8.54	NO	1,313
GIGABIT	171	5	1,773	SPC	1,311
TRAFFIC	411	1,615	13.19	NO	1,279
TECHNOLOGIES	259	329	40.81	NO	1,261
PERFORMANCE	485	2,712	9.27	NO	1,230
COM	315	776	21.04	NO	1,217
SOLUTIONS	279	523	27.65	NO	1,197
PROTOCOL	208	139	77.58	SPC	1,188
DEVICE	298	680	22.72	NO	1,188
SERVICE	730	7,291	5.19	NO	1,181
MPLS	143	0	∞	inf/spc	1,134

Table 2. Keywords in the Dot-com corpus.

The frequency factor is interesting when it is interpreted as typicality for relevance indexes being also essential for term detection, since the frequency of a lexical item in a specific corpus indicates whether its choice is recurrent enough to be regarded as a technical term. Setting *investors* as an example (table 3), its statistical behaviour ranks as the fourth most significant word in the corpus with a score of 2,138 in keyness. Besides, *investors* is rated as technical term in the domain according to the criteria proposed by Chung (2003). As observed in tables 2 and 3, *Term* column reads three possible keys as a result of the ratio value that Chung states to be an indicator of specialty: when a unit is at least 50 times more frequent in CCC or DCC than in *Lacell*, the unit is selected as a term. *SPC* stands for a ratio > 50, *NO* for a ratio < 50 and *inf/spc* means that the ratio is infinite, that is, the unit does not occur in the general corpus and therefore, is deemed a term on quantitative basis. Afterwards, the results must be qualitatively interpreted to discriminate when to categorize proper names like *Obama* (inf/spc) and other low frequency words as technical terms.

WORD	**Freq.**	**Freq. Lacell**	**Ratio**	**Term**	**Keyness**
BANKS	629	1,163	42.89	NO	3,225
MARKETS	524	1,000	41.55	NO	2,660
BANK	655	3,093	16.79	NO	2,351
INVESTORS	392	584	53.23	SPC	2,138
FINANCIAL	608	3,200	15.06	NO	2,070
ECONOMY	494	1,996	19.62	NO	1,903
MARKET	681	5,341	10.11	NO	1,858
CREDIT	431	1,801	18.97	NO	1,635
RATES	468	2,402	15.45	NO	1,613
ECONOMIST	239	148	128.06	SPC	1,585
PRICES	404	1,862	17.20	NO	1,466
BILLION	381	1,569	19.25	NO	1,455
RATE	498	3,652	10.81	NO	1,415
INFLATION	313	972	25.53	NO	1,343
GROWTH	400	2,279	13.91	NO	1,307
DEBT	301	935	25.53	NO	1,291
FED	256	653	31.09	NO	1,181
INVESTMENT	335	1,868	14.22	NO	1,107
MORTGAGE	234	567	32.72	NO	1,099
FUNDS	290	1,252	18.36	NO	1,084
HEDGE	170	166	81.21	SPC	1,029
RECESSION	217	513	33.54	NO	1,028
LEHMAN	128	15	676.73	SPC	1,027
EQUITY	201	380	41.94	NO	1,023
GDP	158	121	103.55	SPC	1,007
OBAMA	112	0	∞	inf/spc	982
FIRMS	252	991	20.16	NO	982
CRISIS	240	1,005	18.93	NO	909
YEAR	867	20,144	3.41	NO	886
CAPITAL	301	2,223	10.73	NO	851
LOANS	176	501	27.85	NO	780
SPENDING	259	1,726	11.90	NO	777
CENTRAL	330	3,313	7.89	NO	763

ITS	1.057	31,536	2.65	NO	736
BONDS	151	357	33.54	NO	715
COMPANIES	324	3,505	7.33	NO	710
ASSETS	164	501	25.95	NO	708
ECONOMIC	327	3,597	7.20	NO	707
MR	605	12,659	3.79	NO	706
INTEREST	383	5,418	5.60	NO	673
GLOBAL	198	1,062	14.78	NO	667
MORTGAGES	110	113	77.19	SPC	658
FUND	222	1,531	11.49	NO	653
BANKING	145	403	28.53	NO	648
QUARTER	229	1,691	10.73	NO	647
ECONOMISTS	117	168	55.22	SPC	644
SHARES	211	1,366	12.24	NO	644
CHINA	194	1,111	13.84	NO	632
LENDING	123	232	42.04	NO	626
GOLDMAN	96	71	107.22	SPC	616

Table 3. Keywords in the Credit Crunch corpus.

Specialized vocabulary is regarded as a cline of lexical units technically loaded, ranging from highly restricted terms to those which share some features with other subject matters. Out of the whole set of keywords in the credit crunch corpus, 173 (14.5%) correspond to specialized terms (SPC), 850 (71%) to non-specialized units (NO) and again 173 (14.5%) coincide with the terms characterized as highly specialized (inf/spc); whereas the percentages increase slightly in the dot-com corpus, with respect to specialized terms (21.5%) and highly specialized terms (24.8%), leaving 53.7% for non-term but relevant keywords (figure 1).

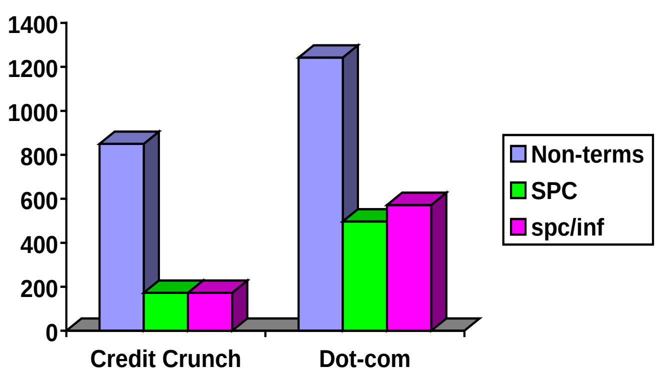

Figure 1. Term classification in CCC and DCC.

Keywords are analysed from a different perspective by bringing Range into play so as to get an overview of the proportion of general words from the main ranges of frequency which take on relevance in each corpus (table 4). Those data may unveil the degree of comprehensibility of the text, since an inordinate amount of off-list words would pose considerable difficulties of cognitive processing.

WORD LIST	DOT-COM CORPUS			CREDIT CRUNCH CORPUS		
	TOKENS/%	TYPES/%	FAMILIES	TOKENS/%	TYPES/%	FAMILIES
One	398/16.39	290/12.91	185	355/28.49	314/26.59	202
Two	133/ 5.48	131/ 5.83	83	97/ 7.78	97/ 8.21	66
Three	292/12.02	287/12.77	166	142/11.40	142/12.02	90
Off-list	1,606/66.12	1,539/68.49	?	652/52.33	628/53.18	358
Total	2429	2247	434	1246	1181	?

Table 4. Keywords in ranges of frequency.

Similarly, it is worth pointing out that general words from the most frequent bands comply with the quantitative conditions established for term detection. This fact points at the need for them to receive a closer qualitative analysis in order to find out whether and why they have activated a specialized meaning in their context of use. Table 5 displays the keywords classified as terms (ratio>50) which fall in the three frequency bands. In the dot-com corpus, two word families (*application*, *provider*) are registered in list 1; three word families materialize in list 2 (*model, phone, shield)*; twelve belong to list 3 (*analyse*, *automate*, *capability*, *conformance*, *device*, *dynamic*, *impact*, *infrastructure*, *network*, *prioritization*, *protocol*, *route*) and 474 types do not belong to any of those lists, among which we could mention the following: *adapter, amplifiers, authentication, backups, bandwidth, blackberry, bottlenecks, broadband, cache, circuitry, coms, configure, customize, dialling, encryption, failover, firewall, Google, handset, interexchange, LAN, laptop, led, lifecycle, micromouse, modulation, nanotechnology, NASDAQ, networld, optoelectronic, outage, radiocommunications, reboot, router, server, switches, Telecoms, telecommuting, telework, troubleshooting, upgrades, vendor, vulnerabilities, waveguide, wideband, wireless, zonealarm*, and successively. The figures in the credit crunch corpus resemble those in the DCC, but the lexical units are clearly distinctive: three families in list 1 (*bank*, *dollar*, *reserve*); three families in list 3 (*economy*, *invest*, *regulate)*; and 173 off-list types, some of which are: *arbitrage, bailout, borrowers, bourses, checkpoints, crunch, deflation, dot-com, downgrade, eurozone, foreclosure, Google, hedge, homeowners, illiquid, imbalances, mispriced, mortgages, payouts, recessions, savers, stockmarket, telecoms, treasuries,* and so on.

WORD LIST	DOT-COM CORPUS			CREDIT CRUNCH CORPUS		
	TOKENS/%	TYPES/%	FAMILIES	TOKENS/%	TYPES/%	FAMILIES
One	34/ 6.34	4/ 0.80	2	20/10.87	11/ 6.36	3
Two	3/ 0.56	3/ 0.60	3	0/ 0.00	0/ 0.00	0
Three	16/ 2.99	16/ 3.22	12	4/ 2.17	4/ 2.31	3
Off-list	483/90.11	474/95.37	?	160/86.96	158/91.33	?
Total	536	497	17	184	173	6

Table 5. Specialized keywords in ranges of frequency.

With respect to the highly specialized keywords (infinite ratio), whose percentages are 24.8% in DCC and 14.5% in CCC, these do not appear in the general corpus and, consequently, they all become off-list types. Regarding the remaining non-term keywords (ratio<50), which do not offer any specialised meaning of their own, but may activate a technical charge in context and diverge from their typical usage, list 1 covers 23.19% and 36%, list 2 shows 10.15% and 11.41%, list 3 reaches 22.06% and 16.24%, while off-list types get 44.44% and 36.35% in DCC and in CCC, respectively.

Once we have displayed the data that characterize each corpus individually, their common features are brought to light as well, with the appearance of the most salient, central and typical lexical units which are shared by both corpora, no matter how peculiar they might seem to each of them. In other words, we are talking about the lexical units which become key in both corpora, jointly and severally (table 6). The dot-com corpus and the credit crunch corpus share no less than 74 keyword families from list 1, 10 from list 2, 36 from list 3, and 38 types that fall out of those lists.

LIST 1: *accord, account, also, back, be, business, centre, company, continue, cost, current, demand, else, employ, exchange, expect, fail, fast, figure, flow, future, gain, grow, help, high, increase, industry, it, large, late, lead, level, look, low, machine, make, manufacture, market, measure, million, month, more, next, not, plan, point, price, product, provide, quarter, rate, real, recent, reduce, report, sale, save, say, sell, September, service, share, small, standard, such, supply, this, value, will, work, world*
LIST 2: *balance, billion, customer, ease, firm, manage, net, pack, scale, telephone*
LIST 3: *acquire, administrate, analyse, benefit, consult, consume, core, corporate, cycle, data, economy, edit emerge, expand, federal, finance, focus, globe, impact, invest, method, output, overall, potential, predict, project, purchase, recover, rely, revenue, secure, strategy, survey, target, trend, volume*
OFF-LIST types: *announced, AOL, asset/s, bankruptcy, bubble, budgets, CISCO, COM, consolidation, DC, default, downturn, executive/s, faculty, forecast, Google, IPO, Kraft, Lehman, leverage, Mack, NASDAQ, offs, OIS, peak, portfolio, robust, Shiller, slowdown, telecoms, transactions, UK, unveiled, updated, USD, website.*

Table 6. Shared keywords in ranges of frequency.

If positive keywords are more frequent and likely to occur in either of the corpora, those which appear in both will reveal the topic or topics common to both areas. Nevertheless, it makes little sense to describe lexical content through individual words in isolation, since words on their own render little meaning, and it is their arrangement in larger units in co-selection which conveys specificity. Nevertheless, and for analytical convenience, the individual word is the starting unit to perform a deeper and further analysis in the lexical combinations occurring in a span of five words to left and five words to right, within its actual context of usage. In our analysis, we have randomly selected ten keywords from each corpus and ten shared keywords from different ranges of frequency, in order to study how the meaning dispersed over 2-word combinations or clusters, which habitually co-occur in the texts, may offer an insightful description of the use of language during the crises. The corresponding results are shown in the next section for the sake of discussion.

4. DISCUSSION: DIFFERENCES AND RECURRENCES IN THE VOCABULARY OF THE CRISES

This study was devised, at its onset, with the aim to find out the potential recurrence level in the lexical fabric of two corpora, or subcorpora, of texts illustrating two different crises, a technological and a financial one, respectively, applying the quantitative parameters which condition term detection. Our final aim would be to detect whether there are lexical concomitances, or informative conformity, in the economic phenomena as illustrated by words in either crisis, or if the contrary was to be true. Constituting, indeed, two specialised corpora –both supposedly within the specialised realms of technology and finance, correspondingly– we initially hypothesised that the differences and recurrences found would be in harmony with the specificity levels of the different lexical fields they belong to (i.e., technology and finance), and that these areas, in turn, would be in equilibrium regarding their own terminological volume. Our quantitative data analysis has evidenced that the hypothesis of obtaining lexical uniformity in both corpora has proved to be untrue. Actually, the first factor that springs to mind in the light of numerical results is that there is a greater terminological specificity in the dot.com corpus than in the credit crunch. This is shown in a variety of factors, the first being the smaller set of types, or word-form variety in the CCC ─for example─ which was rated 2 points lower than in the DCC, thus illustrating a lower lexical burden in the former. Not only this, but also a lower degree of specificity was found in the business subcorpus, the CCC indeed showing a significantly short relevance in terms of keyness. This is revealed, for example, in the fact that the word with the highest degree of keyness in the said corpus, *banks* in its plural form (ostensibly not a technical word but a general one), has a score of 3,225.30, whereas in the DCC, the word that ranks first is *network*, a specific term with 8,991.20 keyness degree. This pattern –less degree of keyness in the CCC, more degree in the DCC– is repeated predictably all through the list of keywords in either subcorpus. Still, the third phenomenon is even less reassuring in relation to our credit crunch corpus, referring to specificity of its lexical phenomena, or terms: indeed, out of the first fifty word results obtained in the CCC, only eight were specific (and not all of them with a high degree of keyness, comparatively with the dot-com corpus), an amount that was more than doubled in the DCC, not counting information-specific words. In numbers, 21.5% specific versus 53.7% non-specific words in DCC, and 14.5% specific versus 71% non-specific words in the CCC. Such terminological phenomena, as analysed quantitatively, may indicate that the present-day economic catastrophe, named the Global Systemic Crisis, is, indeed, a global reality, embracing not only specific, but widespread areas of human activity and cognition and invading its everyday reality. On the contrary, the dot.com crisis was apparently a relatively less noticeable phenomenon, as compared to the economic tsunami we are experimenting at present. Our corpus shows, in fact (and despite the undeniable repercussions it had at many levels) that the fiasco restricted to the technological field and boasted fewer actors at play, having as a context the, then, recent advent of the World Wide Web and a group of Internet-based companies.

If the numerical data obtained have been useful, in terms of detecting the specificity of either corpus, our study aims to go a little beyond these. Qualitatively speaking, we aimed to render how both crises configure the specificity of their verbal output through special and peculiar collocations, and how they coincide at times, in terms of key lexical expressions. With such a goal in mind, we have randomly selected ten words from each subcorpus (some general words, some terms), that have a relative degree of keyness in each in turn, to test how the meaning dispersed over 2-word combinations, or clusters, may offer an insightful description of language usage during each of the crises in sequence. In order to measure lexical coincidence, we also selected ten other words –six general words, four terms– as lexical phenomena that co-occur in both corpora at the same time. Following these premises, our study has developed along the following lines:

1. Specific and non-specific words with the highest keyness degree in either corpus, which behave in a specific way when analysed in clusters, or collocations, along the text, and which may reveal an awareness of each crisis, in turn.
2. Shared words, specific and non-specific, with different degrees of keyness in either corpus. This keyness will be used to establish whether both corpora share a similar content, and that they, subsequently, may hold a powerful relationship between them, related to the crises.

Below, in tables 7, 8 and 9 we display such words, remarking on their quality of terms or not, and their keyness degree. Starting with the DCC corpus, we selected words with a significant frequency and reasonable ratios of specificity.

WORD	TERM	KEYNESS
NETWORK	SPC	8,991.2
SOFTWARE	NO	4,154.8
WIRELESS	SPC	3,722.1
TECHNOLOGY	NO	2,256.9
SYSTEMS	NO	2,145.4
MANAGEMENT	NO	1,787.1
BANDWIDTH	SPC	1,438.3
MOBILE	NO	1,336
PROTOCOL	SPC	1,188.2
SERVICE	NO	1,181.1

Table 7. Keywords analyzed for DCC.

Network is outstandingly the most specific and frequent of them all, developing in clusters such as *network world* (with 270, this also being, notably, one of the most important magazines dealing with communications and the Internet), but also followed by *traffic, management, security, operators* and other collocations that have to do with the realm of knowhow in this area. *Software* is in the same context of technological expertise, being mostly preceded by *management, source, protocol, rational* and *server*, and followed by *development, vendor* and *design*. The same again happens to *technology*, mostly –and predictably– preceded by *information*, but also in clusters having *wireless, clustering, switching, cluster, process* and *spectrum* as qualifiers. Some economic awareness is shown in examples where *technology* is followed by companies and business. *Wireless*, in its turn, is mostly preceded by *LAN* –as the transmission method– but also followed by *access, network, internet, communications, service, connectivity* and others. *Technology* is, unsurprisingly, mostly preceded by *information* again, and is followed by an impressive array of qualifiers in collocations, among which are, but not exclusively, *access, network, internet, communications, LAN*, and other terms and general words of the area. As far as the word *system* goes, *operating, replicating, computer, CISCO, DC, recognition, systems* abound, among many others, the word *system* being scarcely found as a modifier. *Management* is mostly preceded by *network* and *traffic*, being followed by *configuration, policy, web,* and others. The term *bandwidth* is mostly qualified with *more, high, limit, higher, mass, memory*, and followed by *available, required* or *requirements*, and *trading*. In contrast, *mobile* is mainly a modifier, mostly with *phone* as noun, but also *Internet, wireless, network, data, companies* and *devices* as some other examples. In its turn, *protocol* is mainly preceded by *access, internet, datagram*, and *ATM*, and occasionally followed by *software, label* and *stack*. Finally, *service* is mostly followed by *provider* or *providers*, and in a lesser amount preceded by *directory, radio, VPN* (*virtual private network*) and *wireless*.

All in all, the terms refer to a set of connections having to do with state-of-the-art technology and very specialized and innovative communication systems, but little reference is made to economy. On the contrary, it is plain to see how there is a criss-crossing of words (*information, technology, network, wireless, Internet*) that pass back and forth in the corpus, qualifying one another and making the corpus a compact and relatively isolated compilation, in tune with the realm of technology; in essence, a barred, restricted area for non-connoisseurs.

WORD	**TERM**	**KEYNESS**
BANK(S)	NO	5,577.4
MARKET(S)	NO	4,508.2
FUND	NO	1,738.6
EQUITY	NO	1,023.4
CREDIT	NO	1,635.9
MORTGAGE(S)	SPC	1,758.3
ASSET(S)	NO	1,102.1
BOND (S)	NO	7,15.3
INVESTMENT	NO	1,107.5
ECONOMIST(S)	SPC	2,204

Table 8. Keywords analyzed for CCC.

As far as the CCC corpus is concerned, we find a very different panorama: as opposed to the constellation of specificity found in the DCC, the credit crunch corpus is made up of mostly unspecific words, as, we will see, in very specific collocations.

In this corpus, *bank(s)* is the most frequent of words, and even if it is not a term, its collocations make it an undeniably specialized word. It is mostly found as a noun modified by quite a wide-ranging array of expressions, such as *central, investment, big (biggest), reserve, federal, commercial, major, world, royal, national, international, local*, and many others, as the obvious main characters of the Credit Crunch crisis. *Market* is also a strong word in frequency, and its specificity in this corpus is manifest in revealing groups, being mainly modified with qualifiers as widely-ranging as *financial, housing, emerging, stock, mortgage, credit, money, capital, debt, labor* and *bull*, among others. *Fund*, in its turn, appears mostly in the collocation *hedge fund*, or *funds*, as one of the iniquitous products that ostensibly triggered the crisis. It also appears qualified with *monetary, pension, market, bailout* and *investment*. *Equity* also appears in a very revealing set of words, as another infamous main character of the crisis, almost solely as *private equity* (*firms*), the operating companies that are not publicly traded on a stock exchange and whose shady investments are difficult to track, as they are under no obligation to publish their accounts. *Credit* is almost always a modifier, and a very specific one at that, characterizing the crisis itself as a *Credit Crunch*, but also evoking *credit market(s), credit crisis, credit cards, credit risk, credit conditions, credit rating, credit spreads, credit derivatives* and many others. Another very revealing cluster is that of *mortgage*, with *subprime mortgage (crisis)*, as the alternative way to define the onset of the present Systemic Crisis. *Mortgage*, one of the two specific words in the sample, also appears in another sadly renowned collocation, *mortgage-backed* (*securities*), and modifying other names, such as *rate(s), default, market, loan, arrears*, and so on. *Asset*(*s*) is also the main character in the well-known cluster, *asset-backed* (*securities*), as yet another by-product of the crisis, but also qualifies *prices, management, values, markets,* and *sales*, as examples. *Bond yields, bond markets, bond investors, bond prices, bond spreads* and *bond funds* are indicative examples of the specific combinations of an ostensibly unspecific word, *bond*. Again, the most usual collocation for the word *investment* is *bank*, as in *investment bank, bankers*, or *banking*, but the word also qualifies *fund* and *firms*, as instances. Finally, the other specific word in the sample together with *mortgage*, *economist(s)*, appears mostly on its own, or accompanying off-list combinations like *Goldman Sachs*.

Those samples from the CCC give evidence of the need to combine quantitative analysis with qualitative assessment of the researcher. Unspecific words can render very specific meanings when in context, unveiling the essence of the crisis they are expression of: extremely sophisticated products and producers of finance that appear at the forefront of the dramatic economic drying-up we are living.

Our following analysis in over 2-word combinations, or clusters, will be that of ten shared keywords from different ranges of frequency, appearing in both corpora. Hopefully, this will permit us to see whether such corpora share a similar content, and subsequently, whether they hold a powerful relationship between them, or not, related to the crises they are the expression of.

WORD	**TERM**	**KEYNESS IN DCC**	**KEYNESS IN CCC**
NASDAQ	SPC	308.7	95.6
BUBBLE	NO	71.8	385.7
BUDGET	NO	31.4	151.8
COM	NO	1,217.3	53.8
GOOGLE	SPC	139	54.4
TRANSACTION	NO	71.5	88.5
TELECOM	SPC	51.6	122.9
DEFAULT	NO	60	322.2
DOWNTURN	SPC	111.6	287.1
WEBSITE	NO	102.6	34.2

Table 9. Keywords present in both corpora.

The first word under analysis is *Nasdaq*. Predictably, this is a word from the field of economics and should be more frequent in the CCC. It is the DCC, surprisingly, that holds the highest frequency degree, and not irrelevantly, since the *Nasdaq Composite* (the collocation it appears the most in, when not on its own) was the index that peaked in the phenomenon known as the *IT bubble*, during which stock markets in Western countries saw their value increase rapidly from growth in the new Internet sector and related fields, marking the beginning of the dot-com debacle. The next word, *bubble*, is even more revealing, if possible: mirror expressions are found in both corpora, such as *dot-com bubble*, *speculative bubble*, *bubble burst, housing bubble, technology bubble*. It is really eerie to see how these collocations are replicated in both subcorpora, revealing that it was the same phenomenon that made the market crash both times: a surreal increase in market dealing that led to an economic catastrophe, and in both cases –despite the substantial differences between them− appearing in the same fields: technology and housing. As far as *budget* is concerned, its appearance in DCC is almost irrelevant, but it is ubiquitous in CCC, in expressions like *budget deficit* (mostly), *budget report*, or *budget balance*. *Com* is dramatically –and as expected– present in DCC, with a much lesser appearance in CCC, in collocations like *dot-com, com bubble* or *com companies*. *Google* is an off-list word, like *Nasdaq*, but it materializes on its own and shows no collocations in CCC, appearing in DCC with *results*, *search* and *bombing*.

All in all, this analysis has tackled the exploration of our subcorpora in terms of frequency and specificity first, then focusing on the analysis of a selection of words and their clusters in context from each corpus. Finally, we have faced the scrutiny of significant samples shared by both corpora, also in sequences of collocation, reaching some illustrative findings that we lay down in the following conclusion.

5. CONCLUSIONS

This study was laid down, at its inception, so as to aim at the detection of recurrence levels at the lexical layer of two corpora, or subcorpora, of texts illustrating two different crises, the dot.com and the Credit Crunch, respectively, applying quantitative parameters conditioning their character of general words or terms. Our final goal would be to find out whether lexical uniformity and/or word recurrence was taking place consistently in each and every of the corpora under study.

Corpus-based techniques and statistical analysis facilitated the collation and integration of evidence on actual language as deployed in the journalistic genres in either subcorpora, furnishing the analyst with reliable clues to extract those lexical units fairly prone to activate specialized meaning. This quantitative analysis was supplemented with a qualitative study of a significant sample of keywords in context to trace the connection between numerical indexes and specialized meaning. We situated our qualitative analysis along two different lines: specific and non-specific words with the highest keyness degree in either corpus, which behave in a specific way when analysed in clusters, or collocations, along the text, on the one hand, and shared words, specific and not specific, with different degrees of keyness in either corpus. This keyness would be used to establish whether both corpora share a similar content, and whether they, subsequently, hold a powerful relationship between them, related to the crises they are expression of.

The terms found in the DCC corpus pointed to a relatively restricted world of state-of-the-art technology and communication systems, with criss-crossing relations and combinations of the same specific and unspecific words in very specialised combinations and little reference to any crisis whatsoever, for the moment. In contrast, our assessment of CCC indicated a powerful group of collocations that make single terms and non-terms become very specific, when in context and tell the story of the economic disaster in terms of its characters, phenomena and products. The more general nature of the lexical items in the CCC shows that the critical situation is more widespread or more general in the context of the Credit Crunch. But these words in context, in the collocations through which they make their appearance, further reveal that an awareness of such critical situation is much higher in this latter selection. Finally, our group of words shared by both corpora gave us definite conclusions as far as what both corpus have in common, as regards the telling of the crisis: a definitive word such as *bubble* is co-ocurrent in both corpora, and in identical expressions, showing that the phenomenon that pervades the present world economic scenario is by no means a new one. The rest of the co-occurrences are not so illustrating, but nevertheless show that indeed the reality told by our corpora on crises was, at some point, the same.

Our study intends, by no means, to be conclusive. On the contrary, it is an attempt to deal with the lexical world expressed to describe two different financial havocs, and constitutes an effort towards understanding the puzzlingly wide-ranging complexity of lexical phenomena in specific fields. The results are concluding, and could provide an example of how corpus linguistics is a useful tool to unravel similar genres in dissimilar disciplines. We firmly believe that the complexity of the subject should encourage further studies in the field, since corpus analysis is a priceless aid for the lexical researcher as a paradigm of analysis, and may be conducive to enlightening results for the improvement of the area.

NOTES

[1] The Lacell Corpus is a balanced 20 million-word English corpus compiled by the LACELL Research Group at the University of Murcia, Spain.

REFERENCES

Alejo R. & McGinity M. (1997). Terminological Loans and the Teaching/Learning of Technical Vocabulary: The Use of Economic Anglicisms in the Business Classroom. In Piqué, J. and D.J. Viera (Eds), *Applied Languages: Theory and Practice in ESP*. Valencia: Universidad de Valencia.

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

Coxhead, A. (1998). *An academic word list*. E.L.I. Occasional publication number 18, LALS, New Zealand: Victoria University of Wellington.

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

Coxhead, A. (2006). *Essentials of teaching college vocabulary*. Boston: Houghton Mifflin.

Chung, T. (2003). A corpus comparison approach for terminology extraction. *Terminology*, *9*(2), 221-246.

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

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. Arnaud and H. Béjoint (Eds.), *Vocabulary and Applied Linguistics*, (pp.126-132). London: Macmillan.

Mateo, J. (2004). El lenguaje de las Ciencias Económicas. In Enrique Alcaraz Varó, José Mateo & Francisco Yus (Eds.), *Las lenguas profesionales y académicas*, (pp.1277-1290). Barcelona: Ariel.

Nation, P. (1990). *Teaching and Learning Vocabulary*. Newbury House.

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

Nation, I. S. P., & Heatley, A. (2002). Range: A program for the analysis of vocabulary in texts [software]. Retrieved from http://www.victoria.ac.nz/lals/staff/paul-nation/nation.aspx

Nelson, M. (2000). *A corpus-based study of business English and business English teaching materials.* University of Manchester.

Orts-Llopis, M. A. (2005). Neological Patterns in Spanish Legal Discourse; the Phenomenon of Mobbing. *LSP and Professional Communication*, *5*(2), 48-59.

Orts-Llopis, M. A. (2007). La influencia del inglés empresarial en la terminología jurídico-mercantil en español. Paradigmas de incorporación léxica. In C. Balbuena Torezano and Á. García Calderón (Eds.), *Traducción y Mediación Cultural*. *Reflexiones Interdisciplinares* (pp. 339-356). Granada: Atrio.

Orts-Llopis, M. A. & Almela, Á. (2009). Dealing with the inevitable: English loanwords in the translation of Spanish Economic Discourse. In A. Sánchez and P. Cantos (Eds.), *A Survey on Corpus-based Research / Panorama de investigaciones basadas en corpus* (pp. 1070-1090). Murcia: Editum.

Palmer-Silveira, J.C, M. Ruiz Garrido & I. Fortanet. (2008). Facing the Future of Intercultural and International Business Communication (IIBC). In Palmer-Silveira, J.C, M. Ruiz Garrido and I. Fortanet (Eds), *Linguistic Insights* (pp. 9-23). Bern: Peter Lang.

Pickett, D. (1986). *Business English: Falling Between Two Styles*. In *COMLON* 26, 16-21.

Posteguillo, S. & J.C. Palmer (1998). Una Taxonomía Textual del Inglés Científico-técnico y Comercial sobre una Base Léxico-cognitiva. In Juan de Dios Luque Durán y Francisco José Manjón Pozas (Eds.), *Teoría y práctica de la lexicología. IV Jornadas internacionales sobre estudio y enseñanza del léxico* (pp.215-222). Granada: Universidad de Granada.

Rea, C. (2008). *El inglés de las telecomunicaciones: estudio léxico basado en un corpus específico*, (tesis doctoral) Murcia: Servicio de Publicaciones de la Universidad de Murcia. Retrieved from:http://www.tesisenred.net/TDR-0611109-134048/index_cs.html

Rea, C. (2009).Wireless: Some facts and figures from a corpus-driven study. *International Journal of English Studies*, Volumen especial, 91-114.

Scott, M. (1997). PC analysis of key words – and key key words. *System 25*(2), 233-245.

Scott, M. (1998). *WordSmith Tools Manual version 3.0.* Oxford: Oxford University Press.

Sinclair, J. (1991). *Corpus, Concordance and Collocation*. Oxford: Oxford University Press.

Stubbs, M. (2001). *Word and phrases. Corpus studies of lexical semantics*. Oxford: Blackwell.

Tribble, C. (2000). Genres, Keywords, Teaching: towards a pedagogic account of the language of Project Proposals. In L. Burnard, L. & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective: papers from the third international conference on teaching and language corpora.* (Lodz Studies in Language). Hamburg: Peter Lang.

West, M (1953). *A General Service List of English Words.* London: Longman.

