UNIVERSITY OF MURCIA

International Journal of English Studies

IJES

www.um.es/ijes

Wireless: Some Facts and Figures from a Corpus-driven Study

CAMINO REA RIZZO[1]
LACELL RESEARCH GROUP – UNIVERSIDAD DE MURCIA

ABSTRACT
Wireless is the word selected to illustrate a model of analysis designed to determine the specialized character of a lexical unit. *Wireless* belongs to the repertoire of specialized vocabulary automatically extracted from a corpus of telecommunication engineering English (TEC). This paper describes the procedure followed in the analysis which is intended to fulfil a twofold purpose: first, to validate the automatic classification; and second, to gain a better insight on the lexical profile of telecommunication English. The statistical information provided by the variables of frequency, distribution and keyness, are combined with the data extracted from the exploration of the surrounding co-text, in order to describe the sintagmatic relations established.

KEYWORDS: register description, statistical data, specialized vocabulary, lexical associations.

RESUMEN
El término *Wireless* ha sido seleccionado para ilustrar un método de análisis que tiene como fin determinar la naturaleza de la unidad léxica. *Wireless* es un término especializado, extraído automáticamente de un corpus de inglés para telecomunicaciones (TEC). Este trabajo describe el procedimiento seguido para obtener un objetivo doble: primero, validar la clasificación automática; segundo, profundizar en la definición del inglés para las telecomunicaciones. La información estadística obtenida con las variables de frecuencia, distribución y palabras-clave se combina con datos extraídos del análisis del co-texto, con el fin de describir las relaciones sintagmáticas existentes.

PALABRAS CLAVE: descripción de registro, datos estadísticos, vocabulario especializado, asociaciones léxicas.

[1] ***Address for correspondence***: Camino Rea Rizzo. *Dpto. Filología Inglesa. Facultad de Letras. Universidad de Murcia. 30071 MURCIA. Tel: 868 88 86 22. E-mail: camino.rea@um.es / camino.rea@upct.es*

I. INTRODUCTION

Major developments and findings in the field of Corpus Linguistics have boosted research and new advances in other linguistic disciplines. Corpus Linguistics has been notably valuable for the exploration and characterization of specialized languages whose meaningful outcomes may be directly transferable to English for Specific Purposes (ESP) practice.

The recognition of the specific linguistic features and communicative skills of target groups in addition to a commitment to learner's specific needs, are two central concerns of ESP as a language teaching approach. Certainly, Dudley-Evans and St. John (1998) set out three absolute and two variable characteristics intrinsic to ESP. The absolute characteristics refer to the design of ESP so as to meet the learner's particular needs; the use of methodology and activities of the disciplines that it serves; and the focus on the language appropriate to these activities with regard to grammar, lexis, register, study skills, discourse and genre; whereas the variable characteristics relate to the fact that ESP may be designed for specific disciplines and may use a different methodology from that of General English.

On the other hand, specialized languages have been traditionally considered as functional varieties or registers (Biber, 1988; Halliday, 1988), defined in terms of the variation in the recurrence of particular linguistic items in comparison to general language or other registers. Hence, the relevance of quantitative data for the characterization of specialized languages is of paramount importance. As a matter of fact, specialized corpora provide the grounds for register description and corpus-based techniques allow to quantify language features, so that statistical accounts of the language are made available. Furthermore, the quantification of linguistic phenomena presents an enormous potential to distinguish what language items are more likely to occur, since it is thus possible to make statistical inferences about language use.

Both descriptive and teaching approaches seem to converge in one of the earliest published papers on the characteristics of scientific English: *Some measurable characteristics of modern scientific prose* (Barber, 1962). Barber reported on a preliminary study focusing on vocabulary, verb-tenses and subordinate clauses, making use of quantitative criteria and the variables of frequency and distribution. He attempted to obtain a list of words commonly used in scientific and technical English, which could be of interest to students and especially to ESP teachers.

Occurrence probability and distribution are evidence of utility which should influence content choice, sequencing of teaching and time investment in teaching. Nevertheless, as observed by several scholars (Kennedy, 2004; O'keeffe, 2007), there seems to exist a mismatch between linguistic research and pedagogy, and more than three decades of research on corpora have had surprisingly little influence on language curriculum contents. With regard to ESP, the effect has been even less noticeable, particularly on those registers

which have not been analysed so deeply, namely English for telecommunication engineering. Analysis and teaching often merge in the same person: the ESP teacher, who performs the multi-faceted task of an ESP practitioner by conducting need's analysis, designing materials, studying the language and the subject, etc. In brief, the practitioner goes to any length to bridge the gap between what is said by the discourse community and what is taught in class.

In this context, a corpus-driven study on the lexis of telecommunication engineering English was conducted in an endeavour to extract automatically the specialized vocabulary of the discipline (Rea, 2008). The present paper describes the analysis performed in order to check qualitatively whether the statistical classification has been effective. An additional value of the study lies in the amount and type of information obtained on lexical behaviour which contributes to map the lexical profile of the register.

II. BACKGROUND OF THE STUDY

The main research deals with the lexical level because the basic difference between general and specialized language stems from the vocabulary that speakers use for communicating, particularly on the terminology of the discipline. Terminology refers to the group of terms which designate concepts and notions specific to a subject field of human activity. Within terminology, there are both lexical units whose use is restricted to the discipline and units from the general language or other registers which activate a different meaning in the domain. The latter are sometimes considered to be less specialized technical terms or to establish its own category: semi-technical vocabulary. Moreover, the lexical level in specialized languages includes general vocabulary as well as academic vocabulary in academic contexts (Alcaraz, 2000; Cabré, 1993; Nation, 2001; Sager, 1980).

All in all, the lexical repertoire of telecommunications obtained is not a list of technical terms but of the specialized lexical units central and typical of the domain. The list contains the most significant and representative specialized units, according to statistical tests which quantify occurrence probability and representativeness. Specialized vocabulary is therefore considered from a broader perspective, taking the position that it embraces technical vocabulary or terminology and semi-technical vocabulary (Alcaraz, 2000; Hyland, 2007; Nation, 2001). In other words, specialized vocabulary, as a whole, is made of lexical units of different degrees of specialization: both words whose use is restricted to a domain, and those used in other fields or in general language which acquire a specialized meaning in the discipline.

The list derives from the comparison of the general language corpus LACELL[2] (20 million words) with the corpus specialized in Telecommunication Engineering English (TEC), which was compiled for research purposes. TEC is a sample of 5.5 million words of academic and professional written English extracted from a wide range of sources (magazines, books, web pages, journals, brochures, advertisements and technology news), gathered in native and non-native parts of the world and covering 18 subject areas subsumed under seven major areas of knowledge (Electronics; Computing Architecture and Technology; Telematic Engineering; Communication and Signal Theory; Materials Science; Business Management; and System Engineering) and two specializations in Telecommunication Engineering (Communication Networks and Systems; and Communication Planning and Management).

Determining whether a lexical unit belongs to the specialized vocabulary of a discipline is a complex task. In the previous study (Rea, 2008), after testing several methods to identify the different categories of specialized vocabulary (Alcaraz, 2000; Chung, 2003; Farell, 1990; Robinson, 1991; Yang, 1986), the following conditions are established for a lexical unit to be included on the list. First, the occurrence of a content word must be statistically significant in the specialized language in comparison to general language. Then, those keywords are gathered in word families starting from the most significant keyword so as to apply Chung's quantitative criteria on term detection to every family member (Chung, 2003). When all or most members are valued as terms according to Chung's criteria, the family is regarded as specialized. On the contrary, when non-terms outnumber terms, forms are individually treated and considered as specialized independently of the rest of its family. Among the most significant keywords there are families or single forms which are registered on the *Academic Word List* (Coxhead, 2000) or on the *General Service List* (West, 1953), so that they are disregarded unless they are also valued as terms in accordance with Chung. In that case the forms are subjected to a detailed analysis in order to ascertain the cause of such behaviour, since they might have a specialized use in the domain. Finally, our Telecommunication Engineering Word List (TEWL) consists of 402 specialized families plus 1,017 individual specialized forms that amount to 2,747 forms altogether.

III. METHODOLOGY AND DEVELOPMENT OF THE ANALYSIS

The procedure followed in the detailed analysis is intended to fulfil a twofold purpose: first, to check whether the automatic classification has been effective; and second, to describe the lexical behaviour of the sample from TEWL. The involved parameters combine statistical data in relation to frequency, distribution and keyness, with the examination of the surrounding co-text in order to describe the sintagmatic relations established.

[2] LACELL is a 20 million word corpus compiled by the LACELL research group, at the University of Murcia.

In keeping with the comparative approach, the set of empirical and statistical data obtained would contribute to map the lexical profile of this register against the general language: "*systematic differences in the relative use of core linguistic features provide the primary distinguishing characteristics among registers*" (Biber, Conrad and Reppen 1998:136). Subsequently, the meaning of such linguistic items in discourse is interpreted as much for what they express as for what they omit. This conception agrees with Sinclair's first principle of textual interpretation, the open-choice principle: "*This is a way of seeing language text as the result of a very large number of complex choices. At each point where a unit is completed (a word, phrase, or clause), a large range of choice opens up and the only restraint is grammaticalness*" (Sinclair, 1991:109).

On the other hand, sintagmatic lexical relations concern the semantic relationships established between a form and the others that keep company, that is, among words occurring together in close proximity. Those relations are connected to the concept of collocation and to Sinclair' second principle of textual meaning interpretation, the idiom principle*:* "*a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments*" (Sinclair, 1991: 109).

The tools available in WordSmith are suitable for performing the corresponding analysis, as collocates and clusters are instantly retrieved from concordance lines. According to the definitions in WordSmith, collocates are "*the words which occur in the neighbourhood of your search word*" (Scott, 1998). These collocates help to show the meaning and use of the analysed word. With respect to clusters, they are defined as "*words which are found repeatedly in each others' company [which] represent a tighter relationship than collocates*" (Scott, 1998).

Looking into sintagmatic relations leads to pinpoint the prefabricated word combinations used by experts in specialized communicative situations. Therefore, those combinations constitute a characterizing factor of the register as well as an essential asset for producing and understanding specialized knowledge.

Next, the form *wireless* is the example taken from TEWL to illustrate the procedure followed in the analysis, which is structured in four sections: frequency, distribution, collocates and clusters. As mentioned before, the two first sections are related to the open-choice principle, within the framework of a specialized register, whereas the other ones conform to the idiom principle, which imposes the restrictions that open-choice sets free. Once detected the lexical selection in the register and its distribution across the different subdomains, collocates and clusters reveal how vocabulary is employed by the discourse community.

III.1. Frequency

The frequency factor evidences the choice of a lexical item in telecommunications register against general language, and indicates whether such a choice is recurrent enough to regard this item as a technical term. The same type of information is stated for the rest of family members and the label of technical, general or academic family is added as applicable.

Setting *wireless* as an example, its statistical behaviour ranks as the eleventh most significant word in the corpus with a score of 11,454 in keyness (Table 1). Besides, *wireless* is rated as a technical term in the domain according to the criteria proposed by Chung (2003). As observed in table 1, *Term, Chung* 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 TEC 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. All the family of *wire*, represented by *wireless*, is accounted a specialized family because the specialized members outnumber the general ones (6 specialized forms against 5 general forms).

KEYWORD	**Freq. TEC**	**Freq. LACELL**	**Ratio**	**Term, Chung**	**KEYNESS**	**P value**
11. WIRELESS	4,083	171	90.6825603	SPC	11,454.00	0
Related members:	**F. TEC**	**F. LACELL**	**Ratio**	**Term**	**Keyness**	**P value**
WIRED	342	110	11.8079342	NO	622.4	0
WIRE	637	611	3.95948522	NO	554	0
WIRES	288	183	5.97698137	NO	359.5	0
WIRING	209	123	6.45329738	NO	275.3	0
WIRELINE	53	0	infinite	inf/spc	166.2	0
WIRELESSLY	31	7	16.8191544	NO	64.2	0
LIMEWIRE	9	0	infinite	inf/spc	28.2	0
NANOWIRES	8	0	infinite	inf/spc	25.1	0.000001
NANOWIRE	4	0	infinite	inf/spc	12.5	0.000397
HOTWIRE	4	0	infinite	inf/spc	12.5	0.000397
Technical family						

Table 1. *Wire* family III.2. Distribution

The distribution parameter explores the arrangement and recurrence of a lexical unit across the different constituent areas of the corpus. In table 2, the squares on distribution disclose the sections where the lexical unit occurs and where it becomes keyword. The possible *Distribution areas* value ranges from 1 to 9, indicating the number of the area where the word does not occur between brackets and a minus, e.g. (-7). The next square, *Distribution keyword in areas*, reports the sections where the unit is key. Then, it is specified whether the keyword becomes key-keyword, the number of texts, and the proportion that these texts cover in the corpus.

Finally, a graphical representation of word distribution is displayed in a dispersion plot facilitated by WordSmith. The graph shows, for every area, the total figure of forms (*words*), the frequency of the analysed word in the area (*hits*), its occurrence per 1,000 forms and the plot corresponding to those data.

Continuing with our example, *wireless* occurs in eight areas of knowledge but becomes keyword in Signal Processing and in the two specializations (Communication Networks and Systems, 082; Communication Planning and Management, 081). This means that *wireless'* incidence is especially significant in three areas, even though it is present and relates in a lesser or greater extent to all the subdomains in telecommunication except System Engineering. Furthermore, *wireless* is key-keyword in 224 files out of 1,654 which the entire corpus comprises, in other words, the presence of *wireless* is significant in the 13.54% of the corpus.

Distribution areas	**Distribution Keyword in areas**	**Keykeyword**	**Nr of texts**	**Percentage**
8 (-7)	4, 801, 802	WIRELESS	224	13.54%

Dispersion plot

Area	**Words**	**Hits**	**Per 1,000**
081 Esp.Sign	867.208	1.342	1.55
082 Esp.Tele	997.727	1.064	1.07
4 Signal proc	580.936	603	1.04
3 Telematics	1.205.064	831	0.69
6 Business	373.079	109	0.29
1 Electronics	722.823	108	0.15
2 Ar. Comp	329.643	16	0.05
5 Materials	101.241	3	0.03

Table 2. *Wireless* distribution

III.3. Collocation and significant collocates

The concept of collocation refers to frequently occurring contiguous or non-contiguous combinations of words which establish a semantic association, in terms of Sinclair (1991:115): "*words appear to be chosen in pairs or groups and these are not necessarily adjacent*". The strength of association may vary from a certain affinity among words, to the extent that the pattern of association gets fixed and the group of words as a whole develops a meaning to become an idiom.

If the occurrence of two words in near context is so frequent as to notice that their co-occurrence is not due to chance, they constitute a significant collocation. In this respect, collocation has a different value in the description of lexical patterns, depending on the units' frequency and position as node or collocate. Collocates may be either more frequent or less frequent than the node itself, giving rise to upward collocation and downward collocation respectively. Therein lies a systematic difference: "*Upward collocation is the weaker pattern in statistical terms, and the words tend to be elements of grammatical frames, or superordinates. Downward collocation by contrast gives us a semantic analysis* of a word" (Sinclair, 1991:116).

If downward collocation enables the semantic analysis of a word, the recognition of a keyword's significant collocates will contribute to unveil the sense attached to this word in the specialized domain and to clear up the possible doubts about the category it belongs to, either technical, academic or general vocabulary. Accordingly, the collocational pattern of the node is analysed by first finding its collocates and later studying the type of relationship they establish.

Table 3 shows 30 out of the 1,148 collocates that the program displays for *wireless*, from a span of analysis set in 5 words to the node's left and 5 words to its right. Results make it clear how many times and in which position node and collocate co-occur, highlighting in bold the most frequent collocation. Nevertheless, little do they report on an existing attraction or on the likelihood of the co-occurrence. Consequently, the next stage of the analysis concentrates on identifying the node' statistically significant collocates.

N	collocates	total	left	right	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	the	1985	1192	793	223	254	194	144	**377**	0	0	158	234	200	201
2	and	1245	643	602	116	132	121	113	161	0	66	**209**	119	101	107
3	of	1032	798	234	135	103	110	174	**276**	0	0	22	30	75	107
4	a	979	608	371	89	94	87	52	**286**	0	0	41	124	110	96
5	to	963	518	445	**138**	107	91	**138**	44	0	9	132	103	99	102
6	in	752	465	287	73	84	62	**125**	121	0	9	84	71	69	54
7	for	690	503	187	56	76	69	106	**196**	0	6	48	41	38	54
8	is	559	187	372	53	62	41	24	7	0	23	**124**	97	68	60
9	network	409	115	294	20	33	35	25	2	0	**186**	25	31	22	30
10	networks	384	75	309	20	11	13	18	13	0	**157**	65	54	20	13
11	access	322	96	226	24	25	28	14	5	0	**133**	43	9	23	18
12	lan	319	20	299	5	2	4	8	1	0	**266**	11	4	11	7
13	that	303	141	162	27	31	25	30	28	0	0	**53**	39	36	34
14	as	302	163	139	**46**	32	32	31	22	0	0	34	43	29	33
15	communications	297	45	252	12	9	11	7	6	0	**173**	35	17	9	18
16	are	254	96	158	30	34	19	5	8	0	8	**62**	36	26	26
17	with	252	142	110	15	23	19	**48**	37	0	1	32	25	23	29
18	systems	229	54	175	10	9	11	12	12	0	**73**	65	16	11	10
19	on	217	135	82	26	18	20	**40**	31	0	0	12	19	21	30

20	technology	202	50	152	18	8	10	9	**5**	0	**88**	30	13	9	12
21	mobile	201	99	102	14	17	14	19	**35**	0	32	19	17	18	16
22	data	200	67	133	9	17	26	12	3	0	**77**	12	16	11	17
23	by	198	103	95	23	17	**26**	23	14	0	0	**26**	20	**26**	23
24	will	187	61	126	21	16	12	12	0	0	3	36	19	31	**37**
25	services	181	62	119	14	14	14	16	4	0	**46**	39	9	8	17
26	or	178	102	76	23	18	17	**27**	17	0	12	24	13	14	13
27	be	165	78	87	13	**43**	17	2	3	0	0	5	31	31	20
28	equipment	165	26	139	7	6	7	4	2	0	**98**	20	8	8	5
29	devices	155	35	120	10	8	6	8	3	0	**84**	16	9	8	3
30	this	155	54	101	10	16	9	12	7	0	0	25	**32**	24	20

Table 3. Collocates of *wireless*

In statistical terms, significant collocation is defined as "*the probability of one lexical item (the node) co-occurring with another word or phrase within a specified linear distance or span being greater than might be expected from pure chance*" (Oakes, 1998:163). Collocates can be subjected to several tests which allow to quantify this probability and estimate how statistically significant the co-occurrence between node and collocate is. The most appropriate tests for this purpose are MI, Z-score and T-score (Barnbrook, 1996). The first test, Mutual Information, is applied by equation 1:

$$MI = \log_2 \frac{O}{E}$$

Equation 1

being *O* the observed frequency of a collocate in the node's environment, that is, the actual co-occurrence frequency between collocate and node; and *E* the collocate's expected frequency, in other words, the theoretical predicted co-occurrence frequency, calculated as follows:

$$\text{Expected frequency} = \frac{F}{F_{Total}} \times T_{conc}$$

Equation 2

where *F* is the absolute frequency of collocate in the corpus, *Ttotal* is the whole number of tokens in the corpus, and *Tconc* the number of words within the span set for concordance

lines. Let us take *network*, a collocate of *wireless*, to illustrate this operation. The values of *F* (16,649) and *Ttotal* (5,533,705) are known, but *Tconc*'s (3,119) comes from multiplying the number of concordance lines retrieved for *wireless* by 10, which is the sum of the five words to the left and to the right of the node. Then, the expected frequency of *network* is 122.84 (Equation 3).

$$\text{Expected frequency} = \frac{16{,}649}{5{,}533{,}705} \times 40{,}830 = 122.84$$

Equation 3

Once all the needed values are available, they are inserted into the original formula which yields the Mutual Information score for *network*:

$$MI = \log_2 \frac{409}{122.84} = 1.74$$

Equation 4

The higher MI score is, the stronger the affinity or attraction between two words. However, there is a threshold or cut-off value which pinpoints a significant collocate, "*below 3.0 the linkage between node and collocate is likely to be rather tenuous*" (Scott, 1998). Therefore, the attraction that *wireless* exerts on *network* is not strong enough to collocate significantly, since MI score (1.74) is quite below the minimum.

The relationship between node and collocate in the example corresponds to a case of upward collocation, as the absolute frequency of *network* is higher than *wireless*', and this type of collocation does not reflect the node's typical lexical environment. Thus the analysis should focus on downward collocation in order to capture those words whose presence is due to the node's attraction. In agreement with previous research (Almela et al., 2005; Barnbrook, 1996; Jackson, 1988; Nelson, 2000; Sinclair, 1991Scott, 1998), significant collocates are extracted on the following basis:

1. Only collocates whose absolute frequency is lower than node's compute.
2. The frequency of functional words is so high and its co-occurrence so probable that they hardly establish a significant collocation.
3. The observed frequency must be higher than 5 in order to avoid the inclusion of non-relevant co-occurrences.

4. The score given by MI and Z tests must be higher than 3.
5. The score given by T test must be higher than 2.

As far as Z and T scores are concerned, the following formulae are applied:

$$Z = \frac{O - E}{\sigma} \qquad T = \frac{O - E}{\sqrt{O}}$$

Equation 5

where O is the collocate observed frequency, E its expected frequency and σ its standard deviation within the entire corpus. When substituting the variables for the corresponding values, *network*, as a collocate of *wireless*, obtains the following scores:

$$Z = \frac{409 - 122.84}{\sqrt{11.06}} = 25.86 \qquad T = \frac{409 - 122.84}{\sqrt{409}} = 14.15$$

Equation 6

Table 4 shows the 103 collocates of *wireless* which fulfil all the previous requirements, together with the score obtained from the different tests performed.

Nº	Collocates	Z	T	MI	Nº	Collocates	Z	T	MI
1	YHOO	28.31	2.43	7.08	53	SBC	10.36	2.33	4.31
2	VERIO	28.31	2.43	7.08	54	INVESTING	9.33	2.12	4.28
3	CWP	28.31	2.43	7.08	55	RIM	11.47	2.67	4.20
4	BREW	26.17	2.43	6.86	56	APS	12.10	2.84	4.19
5	KERTON	24.45	2.43	6.67	57	AUCTIONS	13.74	3.27	4.14
6	CINGULAR	30.79	3.13	6.60	58	WANS	13.10	3.13	4.13
7	TFRC	61.94	6.70	6.42	59	LICENSE	20.81	5.07	4.08
8	WML	20.78	2.42	6.21	60	SPRINT	9.27	2.30	4.02
9	MULTFRC	38.90	4.73	6.08	61	WLANS	15.10	3.75	4.02
10	BELLSOUTH	25.01	3.11	6.01	62	SMC	15.82	3.97	3.99
11	SMARTPARTNER	17.27	2.20	5.94	63	PDAS	12.09	3.10	3.93
12	WECA	18.35	2.41	5.86	64	ELITECONNECT	15.79	4.07	3.91
13	BACKHAUL	20.99	2.78	5.83	65	MULTIHOP	8.00	2.08	3.88
14	WEBLINKS	16.51	2.20	5.82	66	WEP	15.80	4.16	3.85
15	TELEGRAPHY	16.51	2.20	5.82	67	ADVENT	10.57	2.79	3.84

16	SMCANT	15.83	2.19	5.70	68	BRANDS	9.17	2.46	3.80
17	VOICESTREAM	17.12	2.40	5.67	69	WI	12.13	3.33	3.73
18	ROAMABOUT	27.89	4.15	5.50	70	ADAPTERS	9.51	2.61	3.72
19	MSN	15.65	2.39	5.42	71	NETWORKING	32.43	9.08	3.67
20	OUTDOORS	14.21	2.18	5.40	72	BROADBAND	32.46	9.12	3.66
21	HOMERF	27.65	4.26	5.40	73	AP	11.17	3.18	3.62
22	TRUEPOSITION	17.83	2.76	5.38	74	BARRIER	14.59	4.20	3.59
23	WAP	42.45	6.69	5.33	75	DSSS	8.41	2.43	3.59
24	DBI	13.76	2.18	5.32	76	CORDLESS	8.89	2.59	3.56
25	LANS	69.04	10.94	5.31	77	SEAMLESS	12.06	3.54	3.54
26	HIPERLAN	17.15	2.76	5.28	78	REVOLUTION	9.34	2.74	3.54
27	SIERRA	15.87	2.58	5.25	79	WESTERN	8.22	2.42	3.53
28	MARKUP	15.24	2.57	5.13	80	EQUIPMENT	39.92	11.73	3.53
29	WIRED	50.66	8.83	5.04	81	UBIQUITOUS	9.78	2.89	3.52
30	WPAN	14.67	2.56	5.03	82	DEPLOYING	11.94	3.53	3.51
31	WPANS	12.01	2.16	4.94	83	ARENA	6.78	2.03	3.47
32	MISCELLANEOUS	14.16	2.56	4.94	84	DSL	15.74	4.80	3.43
33	PACIFIC	21.38	3.87	4.93	85	LICENSED	6.59	2.02	3.40
34	LAN	93.21	17.25	4.87	86	TELEPHONY	17.14	5.28	3.40
35	VERIZON	18.54	3.48	4.83	87	FIXED	28.74	8.94	3.37
36	WIRELINE	16.97	3.20	4.81	88	WLAN	20.59	6.44	3.35
37	EXPLOSIVE	13.47	2.55	4.80	89	NETWORK'S	6.99	2.21	3.33
38	ROGERS	16.03	3.05	4.79	90	ALLIANCE	7.98	2.54	3.30
39	HANDHELD	26.27	5.01	4.78	91	HOC	13.48	4.31	3.29
40	VTP	27.56	5.28	4.77	92	DEPLOY	11.45	3.78	3.20
41	NYSE	12.28	2.36	4.76	93	TELEPHONES	7.07	2.35	3.18
42	INDOOR	19.15	3.73	4.72	94	BRIDGES	12.25	4.08	3.18
43	FEDEX	11.67	2.35	4.62	95	OFFICIALS	6.46	2.17	3.14
44	PAGER	14.21	2.88	4.61	96	SENSOR	13.66	4.61	3.14
45	NASDAQ	15.86	3.31	4.52	97	AD	12.92	4.42	3.10
46	UWB	10.16	2.14	4.50	98	PROXIMITY	6.79	2.33	3.08
47	BLUETOOTH	39.74	8.48	4.46	99	POPULARITY	6.29	2.16	3.08
48	WTLS	9.98	2.13	4.45	100	TECHNOLOGIES	29.53	10.17	3.08
49	BLACKBERRY	14.70	3.16	4.43	101	WORLD'S	8.05	2.78	3.06
50	DECT	13.99	3.02	4.43	102	EVOLUTION	13.03	4.56	3.03
51	YAHOO	10.50	2.33	4.35	103	ADVANCES	9.95	3.50	3.01
52	OUTDOOR	14.04	3.15	4.31					

Table 4. Significant collocates of *wireless*

The resulting outcome provides the lexical selection that *wireless* demands to occur in this specialized environment. Those significant collocates seem to be closely related to the concept of *wireless* and participate in constructing its meaning. The definition of *wireless* registered in the Webster Specialized Dictionary (www.websters-online-dictionary.com), reflects some connections among the semantic components of *wireless* and the significant

collocates it attracts; while in the *Cambridge Advanced Learner's Dictionary* (2003), the general usage of the word is shown to be old-fashioned.

> ***General dictionary.***
>
> ***Wireless*** *(noun): (old-fashioned) a radio.*
>
> ***Specialized dictionary.***
>
> ***Wireless*** *(noun): Medium for communication. Transmission by radio waves. An electronic receiver that detects and demodulates and amplifies transmitted signals. A communication system based on broadcasting electromagnetic waves.*
>
> ***Wireless*** *(adj): Having no wires: A wireless security system.*
>
> *Specialty Definition: A term describing a computer network where there is no physical connection (either copper cable or fibre optics) between sender and receiver, but instead they are connected by radio. Applications for wireless networks include multi-party teleconferencing, distributed work sessions, personal digital assistants, and electronic newspapers. They include the transmission of voice, video, images, and data, each traffic type with possibly differing bandwidth and quality-of-service requirements. The wireless network components of a complete source-destination path requires consideration of mobility, hand-off, and varying transmission and bandwidth conditions. The wired/wireless network combination provides a severe bandwidth mismatch, as well as vastly different error conditions. The processing capability of fixed vs. mobile terminals may be expected to differ significantly. This then leads to such issues to be addressed in this environment as admission control, capacity assignment and hand-off control in the wireless domain, flow and error control over the complete end-to-end path, dynamic bandwidth control to accommodate bandwidth mismatch and/or varying processing capability.*

The aforementioned concept of collocation refers to the possible attraction existing among words, but with respect to an individual search form. Nonetheless, individual forms may have different uses and take part of a greater unit of meaning or a recurrent combination of words. Moreover, in a specific register, those combinations are prone to convey a specialized meaning: "*Very frequent words in specialized corpora in fact often tend to aggregate in recurrent chunks to form more specialised meanings*" (Gavioli, 2005:79). Word combinations can be identified by the direct observation of the node in its context, examining the word immediately preceding and succeeding the node, or otherwise by means of cluster's development.

The fact that *wireless* occurs 4,083 times in the corpus entails facing the same number of concordance lines when it is convenient to study the immediate co-text, which means an extremely burdensome task to perform. Therefore, the volume of data is reduced by resorting to the right and left adjacent collocates of higher co-occurrence frequency. The selected pre-node collocates (Table 5) occur with a minimum frequency of 7 (*ghz, oriented*), being 59 (*fixed*) the top frequency. Among the 32 collocates, adjectives outnumber the rest

of categories, followed by nouns and verbs. Abbreviations are included for their noticeable instance preceding a node.

Word class	**Sample and frequency**
Nouns	*generation 44, Bluetooth 41, future 17, speed 13, Verizon 12, safety 10, today's 7, Cingular 7.*
Adjectives	*mobile 35, new 23, optical 20, public 18, pacific 15, secure 15, global 13, available 10, various 9, different 8.*
Verbs	*enable 12, deploy 10, include 7*
Participles	*fixed 59, existing 14*, *using 13, based 11, leading 10, including 8, emerging 7, providing 7, oriented 7.*
Abbreviations	*tfrc 41, ghz 7.*

Table 5. Pre-*wireless* top collocates

The number of post-node collocates (Table 6) increases considerably in comparison to those immediately preceding the node. There are 64 post-node collocates whose frequency ranges from 9 (*nodes, handheld*) up to 266 (*LAN*). In this position, nouns predominate over the other categories.

Word class	**Sample and frequency**
Nouns	*network 186, communications 173/communication 95, access 133, equipment 98, technology 88/technologies 63, devices 84/device 29, data 77, systems 73/system 33, link 56/links 21, internet 47, services 46, application 42/applications 35, broadband 32, sensor 24, channel 22/channels 18, home 22, users 22, Ethernet 21, phone 20/phones 12, connectivity 19, industry 19, client 18, telephony 18, carriers 17, connection 17/connections 10, world 16, location 15, medium 15, solutions 15/solution 11, telephone 15, barrier 14, bridge 14, environment 14, knowledge 14, multimedia 14, backbone 13, evolution 13, interface 12, market 12, telecommunications 12, infrastructure 11, operators 11, routers 11/router 9, transmission 11, standards 10, nodes 9.*
Participles	*networking 43, handheld 9.*
Adjectives	*local 49, personal 10.*
Abbreviations	*LAN 266/LANS 93, WAN 17.*

Table 6. Post-*wireless* top collocates

Once the most frequent adjacent collocates are extracted, a combination of two lexical units is set as a node, consisting of the original node and each adjacent collocate. Then, the corresponding concordance lines are explored in order to infer from the context whether such combinations acquire a specialized meaning. Table 7 presents some concordance lines

where the combination *controller wireless* occurs, providing evidence of the usage and specialized meaning in this register.

*82586 ISO Level 2 Ethernet **Controller Wireless** communicates with the CPU via a shared*
*wireless can be of help. The 82586 **controller wireless** has a "time-domain-reflectometer"*
*functions are implemented by the Ethernet **controller wireless** in conjunction with an Ethernet*
*The receiving structure of the **controller wireless** is similar in some respects to the transmitting*
*access unit. To mate the 82586 **controller wireless** to a transceiver, the 82501 serial interface*

Table 7. Concordance lines of *controller wireless*

Extending the procedure to the rest of pairs, 60 out of the 96 couples have been recognized as specialized combinations. The significant collocates are highlighted in bold with the aim of reflecting their distribution and influence on the closest environment of the node. There are 23 instances of specialized combinations where the collocate precedes the node (***Bluetooth***, ***cingular***, ***deploy***, *digital, emerging, enable, existing,* ***fixed***, *free, future, generation, ghz, global, mobile, new, optical, oriented, public, safety, secure, speed,* ***tfrc***, ***verizon*** + *wireless*), while the specialized combinations made of node and post-collocate reach 45 (*wireless* + *access, application, backbone,* ***barrier***, *bridge,* ***broadband***, *carriers, channels, client, communication, connection, connectivity, data, device, equipment, Ethernet, home, internet,* ***LAN***, *local, phone, router,* ***sensor***, *services, solution, standards, environment, location, market, medium, multimedia, network,* ***networking***, *nodes, operators, personal, system,* ***technology***, *telecommunications, telephone,* ***telephony***, *transmission, users, WAN, world*).

The next step is devoted to the study of combinations consisting of more than two lexical units, in order to complete the lexical profile of this specialized register and evidence the idiomatic use of the language.

III.4. Lexical groups or clusters

A group of lexical units which are repeatedly found together builds a cluster. Unlike collocations, clusters establish a stronger relationship since they are the exact repetition of the same sequence of words. These sequences may have a varying extension, that is to say, there are clusters made of two, three, four or more words, and sometimes, they form embedded structures. Hence, the current analysis starts from the two-word clusters previously identified so as to check whether they belong to longer multi-word units.

Additionally, the concordance lines of the varying-extension clusters are explored in search of collocation patterns. The minimum frequency for a cluster to be computed is set in

3 and the number of units expands from two up to six. The results for *wireless* are as shown in the following table:

Combinations	Volume	Highest frequency
2-cluster	*3,126	50
3-cluster	647	48
4-cluster	250	6
5-cluster	60	4

Table 8. Clusters of *wireless*

The fact that two-word clusters stand for the overwhelming majority is owing to combinations of functional and notional words, and functional words among each other. However, only the groups of content words will be assessed as long as one of them is the node. As a result, two-word clusters are reduced from *3,126 to 352.

Next, the 60 two-word specialized combinations detected before are sought out within the clusters of different units. On the one hand, the most recurrent pattern exhibits a lot of possible combinations resulting from the grammatical system of the language (*in + a + wireless+ collocate: in a wireless system; in + wireless + collocate: in wireless networks; on + a + wireless + collocate: on a wireless LAN).* On the other hand, during the development of clusters, it is noticed that some specialized combinations are included in wider clusters, becoming multi-word lexical units (*Bluetooth wireless technology, wireless networking equipment, wireless local area networks, optical wireless communication systems, third generation wireless network operators*).

When the most frequent adjacent collocates were inspected, the outstanding occurrence of nouns was pointed out. They usually succeed the node and get embedded in an immediately superior cluster (*free wireless backbone, fixed wireless bridges, wireless Ethernet compatibility, wireless location technologies, wireless multimedia communications, wireless telecommunications equipment*). The most recurrent two-word cluster including a noun consists of *wireless+LAN* (226) – even though *LAN* is an acronym from *Local Area Network* which has been nominalized, followed by *wireless+network* (186) and *wireless+communications* (173), which integrate upper-level lexical combinations. Again *wireless LAN* is *the most* prolific combination, taking part mostly in three-word clusters (*wireless LAN vendors, public wireless LAN, wireless LAN chip, speed wireless LAN*) and in six four-word clusters (*wireless LAN data streams, wireless LAN security scheme, wireless LAN security solution, speed wireless LAN standard, generation wireless LAN security, friendly home wireless LAN*); whereas *wireless communications* is less productive but develops four and even five-word clusters (*digital wireless communication systems, optical wireless communication systems, family of wireless communications platforms)*. Concerning

wireless network, it only combines in clusters of three units (*wireless network access, mobile wireless network, available wireless network, wireless network interface).*

A close observation of the immediate right and left co-text of the pairs reveals two behavioural patterns. The first and more common case is found when the two-word cluster behaves as a premodifier of the head of the noun phrase. For example, *wireless LAN* premodifies a great number of nouns such as *access, authentication, chip, data, equipment, market, products, protocols, security, services, standard, systems, technology(ies), transmission, users,* etc. There are many other instances of the same fashion like *wireless network + (access, interface, needs, operators, policy); wireless data + (technology, communications, network); wireless internet + (access, service); wireless telephone + (industry, systems)*.

The typical idiomatic usage of the language in telecommunication engineering is also evidenced by the second type of behaviour. In this case, the head of the noun phrase corresponds to a collocate of *wireless* being premodified by another noun or an adjective at the same time. For example, *(digital, indoor, mobile, optical, safety) + wireless communication(s); (home, public, securing) + wireless LAN; (area, available, fixed, mobile, public, safety) + wireless network; (fixed, oriented, broadband, public) + wireless access;* etc. This pattern allows bringing out two-word specialized units, as those two-word clusters designate concepts and notions specific to telecommunication field.

Finally, the analysis focuses on the adjacent collocates of the remaining two-word specialized combinations and other recurrent two-word clusters (see some examples in table 9). *Wireless'* significant collocates are highlighted in bold to emphasize their proximity and/or inclusion within combinations. It is remarkable how significant collocates are located in one position to the right or left of the combinations. Indeed, the recurrence of those sequences evidences the characteristic lexical behaviour in the specific register, where some of them are associated to take on more specialized meaning such as *third generation wireless LAN, smartpartner pager wireless data service, public safety wireless network, next generation wireless network, wireless networking area networks,* etc.

PRE +	COMBINATION	+ POST
contrast, coolest, non-	***bluetooth** wireless*	*communication(s), device, connection, link, personal, phone, **technology(ies)***
*delivering, spectrum, sells, longer-distance, dubbed, based, service, offering, point-to-point, **broadband**, high-speed, offers*	***fixed** wireless*	*access, **bridges**, **broadband**, bypasses, communications, fills, internet, local, vendors, technology, networks, metropolitan-area, offers, PCS/Cellular, requires, routers, service(s), solution*
third, next-, first-, second, fourth, 3rd	*generation wireless*	*applications, **LAN**, telephone, spectrum, communication, mobile, multimedia, personal, network(s), system(s)*

Ericsson's, testing, developing, finding, support	*new wireless*	*access, applications, border, **broadband**, capacity, devices, gadgets, platform, **telephony**, services, standards, **LAN***
***outdoor**, diffuse, **indoor**, hybrid, includes, multichannel*	*optical wireless*	*applications, architectures, channel, communication(s), system, network*
high, higher	*speed wireless*	*access, connectivity, **equipment**, **LAN**, network(s), provide, transmission*
*bit/sec, **broadband**, Cisco, class, digital, diverse, cost-effective, emerging, mobile, expandable, fibre-, **fixed**, flexible, focuses, oriented, heterogeneous, control-level, LMDS, **multihop**, implement, internet, MAC-level, openwave, **fixed,** fiber-overlapping, provide, public, providing, reliable, shared, speed, standard, real-time, **ubiquitous**, unauthorized, universal*	*wireless access*	*platform, point(s), policy, protocol, increases, PCI, PC, networks, architecture, solution, products, **technologies**, **WAP**, system(s), scenario, designs, using, protocol, **cordless**, channel, layer, technique*
***license**-free, **licensed**, multiple, need, versus*	*wireless backbone*	*links, **equipment***
*customer's, install, typical, intros, focus, **fixed***	*wireless bridge(s)*	*requires, contains, configurations*
*Canopy, **fixed**, full, D-link*	*wireless **broadband***	*channel, designed, products, system(s), solutions, service, applications, access, group, coverage, router(s), channels, industry, technology*
Aironet, enable, disassociate, links	*wireless client(s)*	*associates, **adapters**, users, device, network, store, lose*
*accessing, advanced, **BellSouth**, faster, GPRS, GSM, mobile, offering, **outdoors**, **pager**, proven, provide, robust, secure, specialized, two-way, global*	*wireless data*	*service(s), **LAN**, carriers, network, technology, rate, protocol, solutions, access, subscriber, logging, devices, operates, collection, link(s), communications, **networking**, companies, transmissions, coverage, connections*
***broadband**, different, select, enable, listing, full-duplex, **license**-free, frequency, Ghz, IP, mixing, protocols, LOS, purchase, selecting, specific, testing, evaluate, spectrum, higher-speed, evaluate*	*wireless **equipment***	*typically, family, **brands**, vendor(s), options, operates, solutions, contains, provide, manufactures, models, measures, environmental, features, **outdoors**, software*
***DSSS**, focus, linksys, point-to-multipoint, **roamabout**, simultaneous, performance*	*wireless Ethernet*	*standard, protocol, **bridges**, technology, bandwidth, PC, adapter, card, compatibility*
*enable, **fixed**, GPRS, true, including, Marconi's, PCS, offering, optical providing, secure, sites, **weblinks'**, future, future, **networking***	*wireless internet*	*browser(s), service(s), company, access, mobile, system, protocol(s), applications, providers, computing, easier, capability*
dual-band, standards-based, bit/sec, cellular, certified, client's, low-cost, integrate, effectively, embed, home, enterprise, first-	*wireless **LAN(s)***	*medium, security, hotspots, standard, **WLAN**, services, protocols, architecture, base, access, transceiver, affected,*

*generation, GHz, fast-growing, hubs, include, **indoor**, integrated, manage, maximize, native, new, operate, securing, public, simple, systems, high-speed, support, typical, Cisco, undergraduate, Nortel, **WLAN**, installed, performance, hopping*		*connections, functionality, client, chip, systems, radios, environment, subsystem, saves, transmission, data, implementation, handover, requirements, configuration, **AP**, vendor, market, gear, deployments, application, **technology(ies),** sources, **equipment**, products, users, coverage, solutions, hot, authentication, card, adapter, capability, connection*
*enterprise-class, federal, **fixed**, home, **WLAN**,WLL, **broadband**, GHz, **licensed***	*wireless local*	*area, loop*
*3G, airport, available, area, high-capacity, Cisco's, configure, CPE, default, **DSSS**, half-duplex, editing, external, infrastructure-free, future, generation, install, multi-hop, integrated, local, Nortel, offer, short-range, reduce, RF, safety, entire, high-speed, intelligent, **Cingular**'s, Ricochet, global, public, resilient, heterogeneous, **ad hoc**, spectrum, switched, disparate, next-generation*	*wireless network(s)*	*throughput, node, policy(ies), round-trip, key, standard, operator, module, uses, access, interface, card(s), communications, connections, performance, PC, services, applications, devices, **barrier**, operating, group, setting, subscribers, network, clients, designers, topology, routers, **equipment**s, protection, installations, security, PSEN, infrastructure, **barrier**, means, features, viewpoint, processing*
Bluetooth**, establish, third-generation, **WPAN	*wireless personal*	*digital, area, communication(s)*
low-complexity, efficient	*wireless **sensor***	*network(s), nodes*
*commercial, **deploy**, **fixed**, public, future, GSM,3G, planned, lifestyle-altering, digital, delivering, federal, IP, mobile, offer, **seamless**, running, location-secure*	*wireless services*	*provider(s), offering, needs, universal, performance, revenue, similar, today*
*QAM-based, underwater, international, IP, optical, safety, modern, packet-switching, secure, public, terrestrial, **broadband**, 2G, 3G,4G, compact, CDMA-based, cellular, competing, IPv6,**deploy**, DS/CDMA, microcell, multiuser, NG, statewide, fiber-oriented*	*wireless system(s)*	*architects, transmitting, NGWS, design, standardization, advantages, used, performance, including, 4G, ride, operating, promise, operate, run, destined*
***Bluetooth**, currently, range, digital, cutting-edge, low-cost, exploiting, innovative, commercial, low-power, successful, users, QoS, useless, **fixed**, implemented, network-based, short-range*	*wireless **technology(ies)***	*further, makes, offers, companies, standard, iDen, convergence, platforms, development, designed, present, play, provide, truly, convergence, exist*
establish, existing, 3-G, new	*wireless **telephony***	*services, applications, appear, **auctions***

Table 9. Adjacent collocates of specialized combinations

IV. FURTHER RESULTS AND DISCUSSION

The samples of the language compiled in TEC have been subjected to a series of tests which has allowed the semiautomatic classification of lexical units into three categories: specialized, academic and general vocabulary. The classification has been based on statistical and formal criteria applicable to a vast quantity of linguistic data, thanks to corpus-based techniques. The quantitative statistical results have rendered the clues needed to conduct a qualitative analysis in detail.

The detailed analysis of *wireless* corresponds to the methodology applied to a sample of words from TEWL. The same process has been followed for the analysis of the units representing the word families classified as specialized: *latency*, *impedance*, *Bluetooth*, *firewall*, *protocol*, *chip, bandwidth*, *multicast*, *cache*, *throughput*, *crosstalk, satellite*, *applet*, *cosine, diffraction, netlist, microstrip*, *dialog*, *timeslot* and *unicast*. This set of forms encompasses word families where either all members or most members are automatically valued as terms according to Chung, so that the overall family is regarded as a technical family as well.

The data obtained from the close assessment of the units constitute a source of information essential to ascertain their specialized or non-specialized character, in the context where they activate meaning according to the pragmatic features which define the specific register. The outcome of the analysis has provided empirical evidence that corroborates the specialized character statistically assumed, revealing, at the same time, the typical lexical behaviour of the family's representative.

Every single word combines with other lexical units giving rise to conventional patterns of use which reflect discipline-specific notions and concepts. What is more, some words aggregate into clusters up to five components (*network layer protocol configuration negotiation*, *reliable multicast over satellite networks, next generation satellite system NGSS, transmission line characteristic impedance values*). Similarly, significant collocates often integrate those specialized combinations manifesting characteristic lexical patterns and, on many occasions, distinctive features of the domain when the analysed word does not occur in general language (*optical crosstalk, Spice netlist, analyser applet, shielded suspended microstrip, tree-based reliable multicast*).

The identification of significant collocates is of fundamental importance to disclose the semantic environment related with a word and the sense in which it has been used. Certainly, the significant collocates of a form are usually connected to its definition in the technical dictionary, they contribute to convey its specialized meaning, and also exhibit its specialized use. *Netlist, crosstalk* and *timeslot* are the only exceptions because their frequency in the corpus is not high enough to allow the computation of significant collocates. However, their specialized character is emphasized by the fact that such forms do

not occur in the general corpus, do not distribute evenly across subdomains and, besides, their definition is not registered in the general dictionary but just in the technical one.

All the technical forms analysed share several features. On the one side, they reach the status of key-keyword to a greater or lesser extent, that is, their incidence is significant in comparison to the general language and among the different sections of the corpus, as well as being keywords restricted to the three areas of knowledge. On the other side, they all combine with abbreviations (apart from *latency, impedance* and *applet*) and integrate abbreviations, generating specialised combinations (*CDP: Cisco discovery protocol; WLAN: wireless local area network; MCR: multiple chip rate; DCT: discrete cosine transform; UTD: uniform theory of diffraction*). Abbreviations are a clear sign of the knowledge required to understand this register, thus, the more truncated forms are encountered, the higher the degree of specificity. Likewise, one-member families (*Bluetooth, cosine, crosstalk*) have proved to be highly specialized, like those whose representative is not recorded in the general dictionary (*Bluetooth, multicast, netlist, impedance, crosstalk, microstrip, timeslot*). Their meaning in the general dictionary is usually completely different to the sense registered in the technical dictionary which, in addition, offers a range of uses in several branches of telecommunication.

It is worth stressing the behaviour of *chip* and *satellite*, since they are statistically valued as non-term and represent a family classified overall as specialized, because most members are individually considered as terms. The detailed analysis confirms that both of them are specialized units for the significant collocates found which define the semantic environment. Besides, as far as *chip* is concerned, the senses registered in the general and technical dictionaries are totally different. The meaning and use shown in the technical dictionary agree with the significant collocates encountered in the corpus. Finally, *chip* and *satellite* combine with other lexical units (*flip chip, mobile satellite, chip assembly, satellite operator*), with abbreviations (*LAN chip, NGSO satellite communications, FPGA chip, VSAT satellite technology*) and integrate truncated forms (*VCR: variable chip rate, DBS: direct broadcast satellite, SoC: system-on-a-chip, satellite terminals*).

V. CONCLUSION

The descriptive and teaching approaches on ESP have been herein connected in an attempt to identify students' target language needs by their equating with the most representative specialized vocabulary in the register. This required lexical repertoire (TWEL) was previously detected on quantitative bases, which have been assessed and complemented with qualitative information.

The results obtained from the analysis demonstrate that it has meant to be a useful tool for the purpose of the research. The model of analysis has been effective to check whether

the words statistically classified as specialized units are actually charged with a lexical technical load in the context of use.

The combination and application of the operating parameters have also been helpful to verify the specialized character of those words whose statistical behaviour rated them as specialized units, but belonged to the general or academic vocabulary lists. The detailed analysis has revealed that, in fact, they are used differently in technical texts and have activated a specialized meaning. However, the samples which exemplify such results have not been presented in this paper due to space reasons.

Consequently, we may conclude that the statistical criteria and requirements suggested to detect specialized vocabulary have been validated and therefore, the TEWL would be acceptable as well. Moreover, the outcome supports our standpoint of regarding specialized vocabulary as a set of lexical units technically loaded, ranging from highly restricted terms to those which share some features with other subject matters. Within the coverage of specialized vocabulary, TEWL comprises the most salient, central and typical specialized lexical units of the field, no matter how specifically technical they are.

The so far raw descriptive findings about vocabulary and language patterns are pointless if they are not thoughtfully transferred to teaching. TEWL could be possibly taken into account for teaching purposes, since it may provide some guidelines on vocabulary introduction, sequencing, reinforcement, etc. Certainly, this word list must not be studied in isolation but in context and within the combinations that each unit generates, particularly those which integrate significant collocations, as they all evidence the actual usage of the language in the discourse community.

Nevertheless, further research should be undertaken to instruct the selection and sequencing of specialized vocabulary in teaching materials, in addition to the analysis of all the words from the list so that a more comprehensive description of the register could be offered. Finally, the empirical evidence reported in this paper should contribute to gain a better insight on the specialized lexicon and get a clearer picture of the lexical profile in Telecommunication English.

VI. REFERENCES

Alcaraz, E. (2000). *El inglés profesional y académico*. Madrid: Alianza Editorial.

Almela, R., Cantos, P., Sánchez, A., Sarmiento, R. & Almela, M. (2005). *Frecuencias del español. Diccionario y estudios léxicos y morfológicos*. Madrid: Universitas.

Barber, C.L. (1962). Some Measurable Characteristics of Modern Scientific Prose. In *Contributions to English Syntax and Philology. Gothenburg Studies in English* 14, 21-43. Stockholm: Almquist & Wiksell.

Barnbrook, G. (1996). *Language and Computers. A Practical Introduction to the Computer Analysis of Language*. Edinburgh University.

Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.

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

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

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

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

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

Farrell, P. (1990). "Vocabulary in ESP: a lexical analysis of the English of Electronics and a study of semitechnical vocabulary". *CLCS* occasional; 25. Dublin: Trinity College.

Gavioli, L. (2005). *Exploring corpora for ESP learning*. Studies in corpus Linguistics. John Benjamins.

Halliday, M. (1988). "On the language of physical science". In Ghadessy (Ed.), 162-167.

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

Jackson, H. (1988). *Words and their meaning*. London: Longman.

Kennedy, G. (2004). "The contribution of corpus linguistics to language teaching: Three decades of promise". Paper presented at the *25th Icame Conference*. Verona.

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

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

Oakes, M. P. (1998). *Statistics for Corpus Linguistics*. Edinburgh: Edinburgh University Press.

O'keeffe, A., McCarthy, M. & Carter, R. (2007*). From Corpus to Classroom. Language Use and Language Teaching.* Cambridge: Cambridge University Press.

Pearson, J. (1998). *Terms in context.* Amsterdam: John Benjamins Publishing Company.

Robinson, P. (1991). *ESP Today: A Practitioner's Guide*. Hertfordshire: Prentice may.

Rea, C. (2008). "El repertorio léxico especializado del inglés de la Ingeniería de Telecomunicaciones: Cómo y por qué". *Researching and teaching specialized languages: New contexts, new challenges.* Murcia: Editum.

Sager, Dungworth & McDonald (1980). *English Special languages*. *Principles and practice in science and technology.* Wiesbaden, Brandstetter Verlag KG.

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

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

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

Yang Huizhong (1986). "A New Technique for identifying Scientific/Technical Terms and Describing Science Texts (An Interim Report)." *Literary and Linguistic Computing,* vol. 1:2, 93-103.

