Digital Reading Strategies in Computer English

Alejandro Curado Fuentes
University of Extremadura
acurado@unex.es

ABSTRACT

This paper focuses on the use of web-based discourse by three Spanish undergraduate students in Computer Engineering. Key lexico-grammatical features in Computer English have been analysed by means of corpus linguistics techniques, for which word statistics and collocation functions in the WordSmith concordancer have been highly useful. The information was then used to compare the students' results. The students' reading skills have been evaluated in the documentation process of the three final projects, required prior to their graduation as engineers. Overall observations point to code-switching, re-wording, and key vocabulary identification as strategies that the students demonstrate to be able to operate with significantly after on-line reading. The students also rely more on Internet discourse as a main source of feedback for their projects and useful tool for work. Two of the projects have dealt with the design and use of interactive platforms for language learning (via two different approaches), while the third project has focused on the management of e-learning with the use of Squeak, a special language for programming. The assessment of the three students' reading / interpretation skills has taken place by means of written tasks and personal interviews.

1. Introduction

In specific settings of English use (i.e., ESP-English for Specific Purposes-) in Spain, as happens in most EFL (English as a Foreign Language) countries, a good number of specialised dictionaries and glossaries is designed to address the needs of university students. This is the case of many disciplines, for example, in Computer Science, Collin et al. (2004); in Law, Alcaraz & Hugues (2003); in Library Science, Lozano-Palacios (2002),

etc. Interest in specialised lexicography for translation is also evident in a good deal of Spanish research (e.g., Fuertes Olivera, 2007; Errico & Morelli, 2005; Leitchik & Shelov, 2003; Verdejo-Segura, 2003, etc). The importance of finding effective equivalents may be voiced in Leitchik's and Shelov's words (2003: 90) in that specialised content should be conveniently translated "not only in scientific and technical texts, but as well in publicistic and even art texts". In turn, Cruz Cabanillas et al. (2007) describe the state of computer terminology as often reaching the Spanish context by means of loan words.

Translation of specialised terms is also approached through discourse-level research (e.g., Pisanski-Peterlin, 2005; Alberola-Colomar, 2004; Falcón, 2000). This focus aims to establish translation at a higher plane than the lexical level, i.e., word use and rhetorical structure as combined in text analysis. In this respect, genre / stylistic studies play an important role, as they are increasingly extending to web discourse in academic and professional settings. Specialised lexicography (e.g., Fuertes-Olivera, 2005) is thereby complemented by analyses of web-based discourse, as Posteguillo-Gómez (2005: 5) notices.

With this double-fold perspective-i.e., specialised lexical study and web-based discourse-, this paper describes an approach to Computer English reading comprehension via the digital text (i.e., web-based). I have examined reading procedures in final degree projects that Spanish undergraduate students must accomplish to specialise as Computer engineers, having to produce results (mostly in the form of databases and graphical interface prototypes) based on resources that they have previously decoded and analysed. A follow-up of their web-based resources is contrasted with paper-based material, which the students also had to read. Two writing tasks were given to the students, where they might reflect information processing skills with both types of sources-digital and printed-. In the two compositions, three main strategies have been distinguished, code-switching, re-wording, and key vocabulary identification. In addition, English-to-Spanish translation was observed in the students' writing. A contrastive analysis based on the corpus analysis of key lexicogrammatical features has served as reference for the evaluation of such reading skills or strategies, widely applied during reading performance. In terms of the significant language used, information processing based on printed documents appear to be less effective than on electronic material, according to scores in the tasks and interviews with the students.

2. Strategies and the electronic text

Decoding technical texts written in English is a primary activity in the documenting process of most Computer Engineering projects at University of Extremadura. A general observation is that evolvement occurs in the tasks as a process from decoding to encoding skills (i.e., "from comprehension to production with a diminishing amount of entropy" - López García, 2005: 37). In particular, highly perceived is a type of translation in which there is a high degree of interactivity with the English language via web resources, generally leading to a particular means of interpreting information in which code-switching, re-wording, and key vocabulary identification stand out.

When compared with printed documentation, digital sources seem to be explored more dynamically, with noteworthy use made of skimming and scanning techniques in hypertext links and virtual connectors, in agreement with Posteguillo (2003). This dynamism is closely related to the very nature of electronic documents, which, according to Posteguillo (2003: 31), essentially change the traditional typologies of genres. Factors such as hypertext and digital mode convey a characteristic nature to such texts that undoubtedly influences the way in which the readings are used and decoded. Furthermore, in the case of Computer Science readings, “this dynamic nature is even more acute” (Posteguillo, 2003: 31).

Another significant aspect is the sense of belonging to a group that shares technical knowledge in common areas of Computer Engineering (e.g., expertise in programming, databases, system analysis, etc). Recognition and establishment of a community of computer-knowledgeable readers is thereby established. Virtual environments may foster this growth or reinforcement of grouping, since the importance of physical location decreases, and, instead, as Yus (2005: 83) points out, “the importance of social relationships and network ties” increases. The formation of virtual relationships is significant in the case of students working on their final degree projects, as these students may be working from different physical locations-i.e., they tend to work on final degree projects during their last year of studies, when they may either go home and work from there or have a job that makes them stay away from university during longer periods-. The students access the Internet more as a result, and often check electronic material and explore it; in addition, they increasingly make use of more interactive utilities, such as chat, forums, email, etc, by which they communicate with teachers and peers for academic and informational purposes. All such utilities may be put together in a degree project blended course via a Moodle platform (see bibliography for web reference of AVUEX at our institution, the system where the tools are being used extensively for such aims).

The importance of web sources in the documentation process of Computer Engineering projects is essentially deemed as functional for informational or instructional purposes (e.g. Borja et al., 1999). Characteristically embedded in the academic written register, the type of discourse in such web sources may gradually flow in terms of linguistic-communicative variation within the register, in agreement with Biber (2006: 12): while some texts may deal with methodology more, others are more experimental, offering important illustrations on the use of a given software program or computer algorithm, but not only for research, other purposes may be included (e.g., commercial use, user type, etc). In addition, hypertext dynamism and virtual community are two dominant factors that favour functionality aspects among the users of web sources for final degree projects. In other words, web resources are not only accessed and managed for informational content, but also to examine, retrieve, upload, and update system utilities and tools, among other practical activities.

Along the documentation process, all the students use code-switching at different points and for reasons usually related to content knowledge. This strategy is often used with on-line texts, as some scholars claim, since it generally carries a good deal of “double-voicedness inherent in the emerging genre(s) specific to computer mediated communication” (Tsiplakou et al., 2004: 2). Code-switching tends to be activated when students have to reproduce

significant information that, on the one hand, can be decoded by finding Spanish equivalents, or even by literal translation, and, on the other, may be left unchanged, and English terms are directly handled without being adapted to Spanish rules of grammar and morphology-an important aspect of code-switches, as Sankoff et al. (1990) claim-. This code-switching ability is then applied as computer jargon that would demonstrate the inclusion of the speaker within the computer community (as examined in the data below).

Re-wording, sometimes simultaneously used with code-switching, is a second major strategy being applied. It usually appears more frequently in situations demanding the understanding of new information (previously unknown by the subjects). Re-wording is then activated in the native language, and English is never used. The purpose is mainly to consistently keep heading somewhere, as Sinclair (2004: 69) explains about the dynamic view of discourse-seen as “directional, a succession of changing postures; but it must be heading somewhere”-. The technical English reader is then concerned with informational / expository writing for his / her studies, and tends to use discourse on the web as a strategy to both understand and integrate knowledge in a dynamic process of structural analysis, as Luzón also notices (2005: 141). This re-wording of reported information is especially noteworthy for a subsequent stage of documentation and reinterpreting in the projects.

Finally, key vocabulary is recognised and managed by students both individually and interactively. Words are handled by the students under two conditions, as described by Sinclair (2004: 53):

- a. If there is a prior shared experience of, roughly, definition of the word in the speech community.
- b. If the text structure at the point of using the word allows access to that meaning.

In fact, the lexical items highlighted by the learners in their documentation-i.e., decoding / encoding-processes are elements that satisfy their common goal of using the language for specific purposes; two main categories of lexis, namely procedural and technical, enable the learner to use and interpret the information in this way:

1. Students tend to read procedural items within semantic threads that lead them to assimilation of conceptual information and realisation of goals (e.g., the verb *includes* to examine different salient features included by an application). In the case of argumentative nouns (e.g. Francis, 1986), students interact with meta-discursive elements to come to terms with text cohesion and structure; for instance, the student should find effective correlations between a group of nouns such as *file*, *script*, *applet*, *program*, etc, and the abstract processes being described in the readings-in fact, this abstract process dimension is found as characteristically common in Engineering textbooks (Biber, 2006: 54)-.
2. The identification of technical words mostly corresponds to key concepts, and, as Scott (1997: 3) points out, many topic-related references tend to appear as key in specific texts (e.g., applications, computer processes, etc). These items are also noted down by the

students in the projects, the terms often taking the form of acronyms and abbreviations (e.g., *hub*, *USB*, *XML*, *MySQL*, etc), which are common in computer discourse.

3. The corpus analysis

A corpus has been built with different texts, including those that the students read during their documentation process. The sources are related to the topics chosen for the three degree projects analysed. The digital texts were retrieved from the World Wide Web, as advised by the computer teachers (who, co-operatively and jointly with me, were coordinators of the projects). Serving as reference too, the paper-based sources were printed for the students to read, and also came from recommended web sites. Slaouti (2002: 105) explains that digital and web-based resources enable “evaluation of both product and process as a study skill”. This aspect motivates the form of evaluation described in the next section, started from the corpus analysis.

A focus in one of the three projects is on the design of a web site for interactive language learning (run on a free operating system-GnuLinux-); a second approach is the development of adaptive hypermedia resources for e-learning, and the third project focuses on Squeak programming. While the first two contain, in general, a wider scope in Computer Engineering (demanding knowledge on software development, databases, and networking), the third one is somewhat more restrained to the use and management of graphical interfaces in Squeak (independently of any other systems / applications).

A graphical representation of the corpus sources is included in Figure 1. The corpus contains 30 texts, ten in each topic or project. Two text types dominate the collection, namely journal articles (four in each category) and guidelines or user material (six). While the former tends to describe processes and argumentations for computerised implementations, the latter is generally more direct and instructional, guiding the reader through different steps for work.

Some statistical values are also provided in Figure 1: STTR refers to the number of different words (types) per 1,000 tokens (total number of running words); this score indicates the lexical density of the texts (i.e., the higher the score, the more different words used). As displayed, the Squeak topic presents a lower lexical density (i.e., lower STTR when compared with the other two projects). In addition, because of higher means regarding paragraph length in the readings from the first two projects, structurally, these texts are less compacted, even though they may be more ‘dense’-i.e., with more different words-. Such differences are pertinent to determine degrees of complexity in the texts. Interestingly, as can be examined, in relation to two other values (word length and average sentence length), the Squeak project presents a larger proportion, which means that these texts are more compacted (i.e., longer words are used, and more words are included in the sentences).

Figure 1: Statistical description of some aspects in the corpus

STTR = Standardised type to token ratio / sent. = sentence / par. = paragraph

Aware that the corpus is small-it totals 267,550 words-, I have found this collection of

the three computer topic sources to be useful for the investigation of given features in the sub-language (or specialised language), “reflecting very closely the structuring of the sublanguage’s associated conceptual domain” (McNaught, 1993: 233). As Bowker and Pearson (2002) state, the sizes of special purpose corpora may vary in terms of the type of domain (or sub-domain) investigation. Sub-language description may be allowed by the analysis of three major factors: Ratio between lexical frequency / dispersion, collocational strength, and keyness (e.g. Ooi, 1998: 82-144). With frequency and dispersion, the aim is to distinguish distributed linguistic features across text categories. In collocations, lexical associations and clusters, the items are analysed as primed for use in the sub-language, and in terms of key lexis, each topic / project category can be compared with a larger corpus in order to identify keywords.

The concordance program used, *WordSmith Tools* (Scott, 2000), has generally been run to discriminate common from divergent word behaviour-i.e., generalised (domain-based) word use from emphasised sub-domain features-. Five different wordlists were made as basic reference sources for the data: Three according to each topic category, one large detailed consistency list (DCL), arranged in terms of frequency and distribution of the words across the texts, and one general (frequency-based) word list of the Information Science and Technology (IST) English corpus (Curado Fuentes, 2000). This larger corpus was mainly built, as part of my doctoral research, to locate shared linguistic data among interrelated technical studies in Information Science and Technology; it comprises 857,372 words and a mean of 37,05 words as STTR (similar to the STTR values displayed in Figure 1 above).

The IST word list serves as reference for contrastive analysis. Lexical items are

compared in both lists (the DCL words in contrast with items in IST). Results show that word positions may be quite near-e.g., most prepositions, articles, and conjunctions among the top 25 words (see Table 1)-or may differ (e.g., most content words and some grammatical items, i.e., words further down the lists in Table 1). Obviously enough, the academic written register of computer and information technology texts is corroborated in many similar positions and items drawn from the comparison, but, as going down the wordlists, significant differences begin to appear.

<i>DCL wordlist (target corpus)</i>		<i>IST wordlist (reference corpus)</i>	
<u>Similar positions</u>		Similar positions	
<u>N</u>	<u>Word</u>	<u>N</u>	<u>Word</u>
1	The	1	The
2	Of	2	Of
3	And	3	To
4	To	4	And
5	A	5	A
6	In	6	In
7	Is	7	Is
8	For	8	For
9	That	9	That
10	As	10	Be
11	On	11	As
12	With	12	Are
13	Be	13	On
14	Are	14	This
15	It	15	It
16	This	16	Or
17	By	17	With
18	Or	18	By
19	An	19	An
20	Can	20	From
21	We	21	Not
22	From	22	Which
23	Electronic	23	Can
24	Will	24	Information
25	At	25	At

Different positions		Different positions	
40	But	40	They
41	Hypertext	41	May
42	Communication	42	Has
43	These	43	All
44	Technology	44	Their
45	Page	45	Such
46	University	46	Text
47	All	47	Used
48	Was	48	Also
49	About	49	Its
50	Other	50	Were

Table 1. Contrastive examination of wordlists

Major content words (nouns, adjectives, verbs, adverbs) make the difference in terms of their frequency and distribution throughout the texts. For instance, the adjective *electronic* (# 23) does not occur in the IST list until position 533, while the noun *communication* (# 42) is found at position 1,987 in IST. This type of contrastive observation is also useful to value the important function of content words in the corpus of computer project readings, where they present such high frequency and dispersion values.

Corpus-based analysis with a selection of these content words in the project corpus (e.g., top 500 items in the DCL) should reveal how lexical priming (e.g. Hoey, 2005) with these words takes place within the specialised context-i.e., what type of specific word use is most characteristic (as patterns, semantic preferences, etc)-. This phenomenon tends to occur naturally, as a result of close examinations of collocation frequency and distribution (Hoey, 2005: 128). The concordancer, run on selected DCL items, thereby aims to “make regularities in the language immediately more salient, by collecting dispersed naturally-occurring examples together as concordance lines” (Osborne, 2004: 259).

Resulting collocations and clusters that occur in more than one subject category (i.e., at least two topics) are noted down, as the interest lies here in the compilation of distributed items. Parallel to this notion, a .2 percent cut-off point is established, meaning that at least two text sources must be involved for every 10 concordance lines (2 / 10) examined. Biber et al. (1998: 275) refer to similar cut-scores and explain that proportional values are important “for assessing whether observed patterns are meaningful”. A sample of the analysed collocations and clusters is given in Table 2. A comparison is made with similar IST items in order to check for key target expressions. Asterisks in the items indicate that the given item is different in the opposite corpus, where it may occur below the .2 cut-off point, or even not occur at all.

Target corpus (project texts)	Reference corpus (IST sources)
<i>Examine + number of* / a large amount of</i>	<i>A large amount of / A vast amount*</i>
<i>The world wide web* / on the web</i>	<i>On the web / web-based resources*</i>
Based on / computer-based face-to-face tool*	<i>Based on / Computer-based / based upon*</i>
Related to the	<i>Related to / information-related*</i>
Presented at the conference*	<i>Presented within + time*</i>
To see the accuracy of*	<i>To see / To see further information*</i>
The use of / through the use of	<i>Through the use of / the use of forms*</i>
The development of / development of software	<i>Development of software / web-based development*</i>
Keep / store + as part of*	<i>As part of the research*</i>
Enhance + design of*	<i>Design of / during the design of*</i>
Interact with + interface*	<i>Interact with + computer*</i>
In order to create	<i>In order to create / in order to obtain performance*</i>
Can be submitted / can be submitted by means of*	<i>Can be submitted</i>

Table 2. Contrastive view of concordance-based expressions and collocations in the corpora.

The different collocations and clusters in the analysis can be interpreted as divergent, thereby illustrating particular use according to semantic preferences and even textual collocation. One example is the verb *enhance* and its collocate *design* in the corpus of project readings (see Table 2). This collocation is widely used in the target corpus, appearing with a frequency of 12 times in four different texts (derived forms include *enhanced the design of*, *enhancing its design*, and *enhances the design*). Of the 12 instances, there is only one in the passive voice (*design was enhanced*), which predicts that the colligation of active voice with *enhance + design* in the project texts is highly probable (2.8 times more than random occurrences would have predicted). This pattern may also be considered as a textual collocation (e.g. Hoey, 2005: 125) because the use of this pattern is made within the sub-domain of computer graphics, semantically differentiated from other technical contexts, as Table 2 shows for the reference corpus (i.e., IST). In particular, four texts in the Hypermedia and Squeak projects use the collocation when describing computer graphics for the design of multimedia interfaces, a context in which the lexical item is thus “primed” (if I may borrow Hoey’s terminology).

By and large, specific lexical-textual correspondence can be easily detected via this corpus analysis in the case of those collocates differing from one corpus to another (Table 2). Of course, the corpus of project readings is not meant to satisfy general claims about computer discourse-because of, obviously enough, its small size and restricted focus-but, on the contrary, as McCarthy (1998: 151) states, the corpus should help us “to have some idea of what sort of thing one is looking for in order to use the power of the computer[-based analysis] most efficiently.” In other words, the lexical data serves as good indicators of the type of key language to be examined and exploited in context for ESP or EAP (English for Academic Purposes) work.

Other cases in the analysis, in contrast, include lexical items that appear in the two corpora with a greater affinity in their use. An example is the cluster *in order to create*, widely employed to denote purposeful actions in database texts. As more rhetorical-functional items, these clusters tend to indicate functions such as classifying, defining, exemplifying, and so on. They are likewise common in the project corpus.

With the closer topic-based inspection of the target corpus, every topic-based frequency wordlist is checked in comparison with the IST (reference) wordlist. A resulting list of keywords (Table 3) displays pivotal content words within their corresponding project texts. Key associations, identified with other words, lead to the observation of variation in terms of associated content knowledge. Most items are technical, referring to concepts found in each topic category (e.g., *Smalltalk language* in the Squeak corpus). Some items are found to have clearly denoted semantic preferences in relation to activities or procedures conducted in that set of texts (e.g., *electronic publishing* or *e-learning* in the Interactive web category).

<u>Interactive web</u>	<u>Hypermedia</u>	<u>Squeak</u>
<i>CMC environment</i>	<i>Applied algorithm</i>	<i>Squeak-based</i>
<i>Hypertext writing</i>	<i>SAT-K problems</i>	<i>Project feature</i>
<i>Electronic publishing</i>	<i>Streaming media</i>	<i>Smalltalk language</i>
<i>E-learning</i>	<i>MySQL DB query</i>	<i>Non-relational DB</i>
<i>E-journals</i>	<i>Variational bounds</i>	<i>Image run</i>

Table 3. Keyword-based analysis of three topics in contrast with the reference (IST) corpus

4. Reading performance results

The three project students read the different project-related sources before the experiment took place. For the virtual (on-line) course texts, three sources (taken from the target corpus) were made available within each topic category. Students accessed and viewed these sources via AVUEX, the university's interactive web-based system that enables the instructor / administrator to keep track of the students' activities (e.g., by checking number of links made to the sources, views, time spent reading, etc). In this scope, as Slaouti (2002) claims, interactivity in the system makes evaluation of both product and process possible.

In the case of printed texts, students also had access to paper-based readings from their respective project categories. The three students carried out their reading at home or at the computer lab during their own free time and for a period of slightly over a month. Since all the material was handed out and made available at the same time, it was up to the students to decide which sources to examine first (three for each reading set). After the deadline for reading completion, the evaluation took place and form by means of two five-page reports, one for each set of texts. Also, interviews were conducted with the students once their reports had been scored.

Table 4. Linguistic items used by students for reading / interpretation of texts.

<u>On-line readings</u>			<u>Printed texts</u>		
Interactive web	Hypermedia	Squeak	Interactive web	Hypermedia	Squeak
CODE-SWITCHES					
<i>E-learning</i> (e.g., beneficia e-learning) <i>e- journal</i> (e.g., con e- journal arriba) <i>real time</i> (e.g., con programa real-time)	<i>head-initial</i> (e.g., marking con head-initial) <i>marking node</i> (e.g., especifica node y parsing de node) <i>parsing web form</i> (e.g., envía web form)	<i>Database</i> (e.g., con la database) <i>project</i> (e.g., funciona como project) <i>set up</i> (e.g., instala set-up)	<i>face-to-face</i> (e.g., conferencia face-to-face) <i>e- journal</i> (e.g., ven un e-journal) <i>set up</i> (e.g., hace set up)	<i>Attachment</i> (e.g., trabaja con attachment) <i>marking</i> (e.g., marking de página hipertexto) <i>surrogate</i> (e.g., vector surrogate)	<i>Database</i> (e.g., consulta a database) <i>set up</i> (e.g., con set-up) <i>submit</i> (e.g., función submit para)
RE-WORDING					
<i>Enhance + design of</i> = para ayuda del diseño gráfico <i>Number of conflicts in this system</i> = presenta cosas a resolver <i>We are unable to</i> = no se puede calcular en este caso	<i>in the process of + running</i> = durante el muestreo <i>Decision made that this should</i> = que mejor se opera con <i>Number of conflicts</i> = número de diferentes inconvenientes si no se indica	<i>in the process of + running</i> = invoca items en el proceso de <i>Set up + needs to be able to</i> = se instala setup y ya tiene que visualizarlo <i>In real time</i> = conectado simultánea-mente	<i>Operate + in terms of + interface</i> = que se muestra en el interfaz gráfico <i>Number of conflicts in</i> = importante para buscar soluciones <i>We should be able to</i> = pueden conseguir que funcione en tiempo real	<i>Related to the current term</i> = que se relaciona del mismo modo con este término <i>It is easy to extend</i> = se puede extender esa sintaxis <i>Meaning derived from + components</i> = se deduce más si se incluyen más componentes	<i>Operate in tandem</i> = funciona simultánea-mente con el otro sistema <i>Underspecify interpretation</i> = genera dudas si no se especifica el nombre <i>With streaming video</i> = con elementos multimedia
KEYWORDS					
Procedural:					
<i>defined in terms of models such as to note that</i>	<i>Categories such as Described as Specified as</i>	<i>defined as saved as state that</i>	<i>and so on Illustrate To include</i>	<i>Specify that To realise To save</i>	<i>defined as specifying translated as</i>
Argumentative:					
<i>Click to See Figure # See the results in</i>	<i>Table # To find the solution In (#)</i>	<i>Back to top To find Figure # See the data in</i>	<i>For further reading Representation of Examples are</i>	<i>example of See the analysis in To find the solution</i>	<i>See Appendix See the example in The instance of</i>
Technical:					
<i>E-learning Cookies Xterm</i>	<i>Streaming MySQL Surrogate</i>	<i>Project Real time Squeak</i>	<i>Script Semantic Web editor</i>	<i>Algorithm MySQL network</i>	<i>Squeak SmallTalk Video (stream)</i>

The reports asked about the management of computer-based resources and media for teaching / learning purposes, a common goal in the three projects, as such work was coordinated by both computer and language teachers. The final prototypes in the projects would also have to demonstrate having coped with the technical challenge derived from the design of an administrator-based information system (either interactively or by means of software and adaptive hypermedia-i.e., multimedia and hyperlinks-). As a result, understanding the text sources is directly related to academic / technical competence through the achievement of multidisciplinary solutions (e.g., combining computer knowledge, graphics, design, pedagogy and foreign languages).

The five-page reports must address competence issues by answering two main questions: 1) What concepts are discussed in the texts and key information to be known for the project; and 2) What specific language has been decoded to be able to understand such concepts, making sure that the different expressions and terms are explained. The two reports had to be given to the instructor at the same time; therefore, it was the student's choice to decide which report to write first. The students were given the option to write the reports in Spanish if they felt that in this way, they could explain the concepts more effectively; consequently, all three students preferred to use their native language in the tasks.

In Table 4, a synthesised collection of the data from the compositions has been provided. The two documentation formats-printed and electronic-constitute headings; also, on this table, the answers given by the students are categorised under the three main strategies used: code-switching, re-wording and key vocabulary. The students were never told about the importance of using such skills, and much less, according to the type of medium-printed or digital-in which the sources were accessed and read. Instructions were merely given about the importance of understanding the concepts, of which they should make sense to describe their relevance to the projects. In Table 4, students' decoding comments about the specific items are listed after their retrieval from the reports (all written in Spanish). These answers are included in the 'code-switches' and 're-wording' sections. In the case of 'keyword identification', these are items explicitly mentioned by students as lexical markers of crucial information. For all categories, only representative examples have been selected.

The students referred to explicit lexical-grammatical items from the texts. The examples of code-switching in Table 4 first present the original word and then how it was used in the compositions. The re-wording items also first include the original phrases (mostly corresponding to significant corpus-based language) and then the re-phrasing of such items as literally written by the students. In the vocabulary categories, only English words are shown (the Spanish equivalents used have been omitted, and in the case of some technical words, matches in Spanish are unfound-e.g., *cookies*, *xterm*, *streaming*, *Squeak*, etc-).

There are some items appearing in more than one category (e.g., *in the process of + running* as a reworded phrase in two different projects). Repetitions have also been noted down when two different reports by the same student include them (e.g., the code-switch

e-journal, or the reworded phrase *number of conflicts in* within the Interactive web project). Most coincidences are given in the category of argumentative items, with around 20 percent shared. 12 percent of procedural items are also common, and 10 percent of code-switches appear in more than one project composition. In terms of the number of repeated items in both digital and paper-based reports, 18 percent of repetitions are given in the Interactive project compositions, compared to 14 percent (in Hypermedia) and 9 percent (Squeak). Overall, and in terms of quantitative measurements, as deduced from the results, reading in both digital and printed form elicits a good deal of linguistic devices related to code-switching, re-wording and key vocabulary identification via the compositions.

Once the data has been collected and processed in this way, findings from the corpus-based analysis above are used to compare and evaluate achievements. Based on the lexico-grammatical and semantic significance of the collocations, the linguistic items used in the decoding strategies are grouped according to three evaluation levels or degrees: First, if the relevant corpus items are explicitly brought to attention and realised in the compositions by means of code-switching, re-wording and / or key vocabulary-i.e., those items that tend to occur with a .2 cut-off score or more, or are keywords in the target corpus-, each item is rewarded five points. Secondly, if literal translation is used in the tasks, the linguistic structures get one point each. Thirdly, if the translation is done wrongly-i.e., has mistakes based on lexis and grammar-, a value of -1 is deducted from the scores.

Table 5 displays the results derived from this type of evaluation. In this table, the three strategies have been again used as reference for reading comprehension and processing assessment. No items used in any other strategy (e.g., summarising or stating opinions) have been contemplated, except for translation, a skill also found to develop with corpus items. The focus is thus placed on the use of corpus language in the given strategies. In the search for qualitative differences between printed and digital text-based readings, the various items in Table 5 are scored according to their degree of effectiveness, measured in terms of corpus-based significance.

	Interactive web		Hypermedia		Squeak	
	<i>DIGITAL</i>	<i>PRINTED</i>	<i>DIGITAL</i>	<i>PRINTED</i>	<i>DIGITAL</i>	<i>PRINTED</i>
<i>Code-switching</i>	35	25	30	25	30	35
<i>Re-wording</i>	80	60	90	55	45	40

<i>Vocabulary</i>	90	75	100	85	85	50
<i>Translating</i>	8	6 - 2	10	9	8 - 1	8 -3
TOTAL SCORE	213	164	230	174	167	130

Table 5. Scores obtained by students in strategy use.

As reflected by the total scores, a wider use of all the strategies with corpus-based language is made in all the projects when digital texts have been read. Only with code-switching in the Squeak project readings is this fact contradicted. There are some higher scores in the Hypermedia project, and lower in Squeak. These two aspects coincide with the fact that the texts were longer and shorter respectively (in the case of the Squeak project, the electronic texts were a bit shorter than the printed ones).

Drawn from these findings, two major hypotheses may be that: 1. the medium in which the texts were read will affect what and how linguistic items are used within the same topic (hypothesis 1-i.e., h1-), and 2. specialised content language is to be used for decoding purposes according to the topic of each project (hypothesis 2-h2-). To test these further, the results from the linguistic evaluation (Table 5) are contrasted with the students' own observations and opinions, described in the next section.

5. The interviews

Given the results (Tables 4 and 5) thus far, important information on content items and scores has been obtained and examined. Task management has been checked as a product by looking at what students did in terms of their text interpretation skills. As a second step, three separate 15-minute interviews with the students have been done so that a focus on students' work may bring the learning process to the fore; using the oral and technological means available-i.e., a "practicality scope", as Chapelle (2001: 52-92) states-is done by probing students' task development. In addition, in the case of the digital texts, tracking has been done regarding the time taken to read the sources, number of hits and views performed, links made, etc.

In terms of what sources were read first, all three students admitted to reading the printed texts first: at home or at the school library during the first and second weeks. In all cases, they stated that their main means of approaching printed documents was by highlighting and noting down notions in Spanish (often translating). In turn, they occasionally took a look at the online texts (the student in the Hypermedia project did so

more often, even in school during breaks, because of his use of a networked laptop computer-this information does not contradict recorded activities during this time in the Moodle system-).

When they dealt with the digital sources, the AVUEx course in Moodle demonstrates that the largest number of visits to the online material took place at the end of the fourth week, with an average of five to eight visits a day during a six-day period (lasting an average of 10 to 15 minutes per visit). Links from these texts were also activated at this stage, but in addition, during the first three days, the students in the Hypermedia and Squeak projects were already clicking on some links. As said in the interviews, they performed links to different programs and utilities (e.g., Flash and multimedia programs, Apache systems, Squeak database, php-nuke platforms, etc), and, in some cases, they downloaded the programs in order to try out the software; in others, the purpose was to check complementary information. During the fourth week, the links served to explore further information.

When asked about which report they wrote first, two students (Hypermedia and Squeak) said that they had already finished the printed source-based assignment before they did the online work. In turn, the Interactive web student said he re-read the paper-based material, which he first approached a few weeks earlier, at the end of the one-month period. When asked why, he answered that he felt he had more information after the completion of the online assignment. In this respect, there seems to be no apparent relationship between the length of text / number of words in the projects, and preferred order for carrying out writing (as stated above, the Hypermedia and Squeak texts being respectively longest and shortest of all three).

In relation to the use of code-switches, the students coincide in their emphasis of such terms because they are important for the subjects (a fact corroborated with some of their used keywords). The Squeak student also argued that it is easier to visualise the application and its functions by using English words directly.

As seen in Table 5, idiosyncratic language-i.e., specialised content-related-is especially noticeable in the on-line material reports. Students gave two main reasons for why they may have used more code-switches and key vocabulary in the digital text-based tasks: 1. they could activate web links and headings (visual aids) that explained terms more extensively and 2. the linked resources contributed to locating the specific terms more easily, which allowed for more examples and contrasted information. These ideas seem to fit in with the notion of a digital community integrated by computer knowledgeable readers who realise their membership in the use of specialised content language, interpreted from inter-related web sources.

The information gathered in Tables 4 and 5 above is also confirmed to some extent in the case of the Squeak student's comments. While the two other students speak about the use of terminology such as *e-learning*, *e-publishing*, and *marking*, the Squeak student focuses on different words, like *Squeak* and *Smalltalk*. In the first two cases, the students tend to use terms in code-switching, easily understood by the computer community and even CALL (Computer Assisted Language Learning) people-e.g., the English language

instructor-. In the case of the Squeak student, he is not using code-switches at this more generalised level, and, instead, the names of the computer applications and processes are more specific, functioning as keywords and technical matter, but not as convergent language within the community (e.g., the other two students had never heard of things like a *project* in this context of *Smalltalk*).

Nonetheless, all three students agree with the fact that their topics entail knowledge of specialised language. The use of significant items that come from the corpus analysis, based on frequency and distribution in the project corpus, is a direct demonstration of the students' perceptions. For example, the Interactive web student said that a more restricted focus can be derived from his texts, for which he acknowledges the need to know vocabulary in the field of graphical design. He thus sees this field as a "defined technological arena", as a code related to his own project development, distinguishable from other works. Concerning this aspect, as examined in the conclusions below, the corpus-based exploration of the contents and the students' views of such material tend to match.

6. Conclusions

The three main features or reading strategies used by the three Computer Engineering students for their final degree project documentation-code-switching, re-wording and key vocabulary identification-, have been described above. Such features have been activated in both digital and paper-based material readership. Table 4 displays different examples of linguistic items used in such strategies, often coinciding across project categories and reading media. As chief objective in this paper, the efficient activation of such features for learning purposes has been evaluated by looking at the use of significant corpus-based language. In this regard, the contrastive study (Table 5) has enabled the observation of code-switching, re-wording, and key vocabulary recognition with more corpus-based items in the particular case of digitally accessed documents. The specific lexico-grammatical items derived from corpus analysis constitute reference data with which to measure up the students' perceptions of what items are significant in the texts for their projects.

In addition, as Table 5 also shows, translation is an important skill, carried out with both digital and paper-based material, albeit with some more mistakes in the second type of format. Regarding this difference, no additional feedback from the interviews with the students has been obtained, and no reasons for this divergence can be detected in relation to document format, mainly because the mistakes were common lexico-grammatical deviations in Spanish university students that have basic-to-intermediate English levels (e.g., modality was translated as obligation in *the user may log on*, or the conditional was not identified in the items *as long as* and *provided that*-).

As a consequence, there appears to be a relationship between the increased use of the three targeted skills or strategies with significant corpus-based items and digital academic text (journal article and guidelines) readership, as examined. Also, the students' general

impressions on the favourable influence of the web-based electronic medium on document reading and comprehension tend to indicate positive observations in the direction of project development for Computer Engineering. Code-switching seems to play an important role to bring concepts under the looking-glass, and in the Interactive web and Hypermedia projects, as values in Table 5 illustrate, such a strategy is used more with electronic texts. This relationship may be verified by the Hypermedia student, who stated in his interview that, for his project, he needs to use English words referring to concepts (e.g., references in hypertext code marking). For specialised discourse, this need to deal with “double voicedness” (Tsiplakou et al., 2004: 2) in the texts is thus invoked by the student without his being actually aware of the process. In so doing, he is providing positive feedback for hypothesis 2 (h2) above-i.e., that specialised content language is to be used for decoding purposes according to the topic of each project-.

The Squeak project student also commented on this obvious need to refer to untranslatable terms seen in the readings, but his scores in this category were higher with the use of printed material than with digital sources. In his case, the more dynamic nature of electronic text did not correspond with the production of more code-switches, while the writing post-task based on the printed material captured more examples of such terms.

In the re-wording category, all three project virtual readings seemed to prompt a higher reliance of this strategy on corpus-based items. Increased use of re-wording with the digital material also seems to contribute to improving the interpretation process. A quantitative view of the items is clearly observed (i.e., effective re-wording or explaining seems to take place with all types of material, but the number of re-wording instances is larger with digital sources). In the use of keywords, such a divergence between on-line and paper-based material is even more acute, as revealed by the differences in scores. In addition to the medium used, there is a distinguishable sense of specialised language recognition, particularly strong with the items (procedural, argumentative and technical) selected as key. Students in fact claim that the key concepts were checked by definitions and explanations, complemented by examples and visual illustrations. Often, this specialised content could be graphically examined and explored on the Internet in order to be understood effectively.

In this sense, many linguistic items from the corpus analysis (see Tables 2 and 3) could be easily identified by the students. One example is that all three students could point to argumentative nouns as significant, describing such items in graphs and charts displayed by digital means that the printed sources lacked. A parallel inspection of procedural words in Table 4 seems to demonstrate that some significant meta-discursive items can be identified and recalled by the students, who often used these keywords as part of the re-worded expressions. The importance of this semantic area of procedural items is generalised in all the projects. The following excerpt, taken from the Hypermedia task, may serve to illustrate the use of re-wording and subsequent realization of the procedural data within the phrases:

La estructura especificada como PHP y MySql (acompañado generalmente del servidor web Apache conectado multi-user), también está integrado con (...) (...) y las palabras specified as y described para referir el tipo de herramienta especificada (...)

The structure specified as PhP and MySql (generally accompanied by the Apache web server connected as "multi-user"), is also integrated with (...) (...) and the words "specified as" and "described" to refer to the specified type of tool (...)
(My translation).

Examples like this may lead to a deduction similar to Sinclair's (2004: 53): That key lexical information is recognised and exploited by the students when there has been a previous shared experience of the concept in that community of users, and, when the "text structure at the point of using the word allows access to that meaning" (Sinclair, 2004: 53). In the excerpt in Spanish above, a typographical proof is the student's integration of English terms without the use of inverted commas.

The examples of argumentative words specified in the tasks also enable similar conclusions on the use of specialised content language (hypothesis 2-i.e., h2-) and electronic media (h1), thereby providing further evidence in favour: There are in fact many more argumentative nouns, even repeating across the three projects, in the compositions based on the students' use of electronic resources. Words such as *file*, *form*, *template*, etc, appear to be fluently manageable in the students' project reports, i.e., these lexical items are more likely to show up in the content discussions after working with digital material.

The general scope derived from the reports, evaluation scores and interview comments is positive for the formulation of the two hypotheses (h1 and h2) above: Good performance is perceived in terms of specialised content language (h2), and this competence is higher when digital resources have been managed (h1). There are minor exceptions in the case of the Squeak project student, who has provided more examples of code-switching with corpus-based items in the processing of paper-based material, or in the case of literal translation, where a relationship with the electronic format cannot be easily suggested. In such cases, nonetheless, a tool like the text statistics (Figure 1) for the Squeak category may be used for further interpretation of the results. As observed above, in the Squeak project texts, there are higher sentence and word lengths, a fact that may imply less conciseness and, in turn, more structural complexity. Also, the Squeak sources refer to content knowledge in a narrow sub-domain or subject, uncommon out of such texts. In any case, variation in the performance results reinforces the idea of further testing, which should be obviously done at a larger scale in order to offer a more consolidated evaluation of online discourse reading; in other words, if corroborated by a larger number of case studies, the productive relationship between the digital medium and reading skills opens up highly important and necessary investigation paths into language use within academic contexts.

Such statements about web-based reading performance and ESP, based on the findings from the experiment with the three students, do not mean to be conclusive; instead, they should build up naturally in their contrastive study with further web based reading testing

and linguistic-communicative competence analysis. The observation of the three reading strategies with corpus-based items serves to present empirical evidence, via their evaluation, towards reading comprehension assessment in a foreign language for specific purposes. Academic literacy, technical competency and foreign language command would thereby find a natural bonding in this scope of ESP case study. Text decoding techniques seem to receive a positive influence from the digital medium in terms of performance; this command would benefit not only the students' reading achievements but also their own final degree project perception and design. Resources being digitally accessed seem to benefit autonomous learning, as concepts are approached and exploited with little teacher intervention. The students' comments corroborate these views in that they see that working with specialised online discourse can function as a crucial step in the final project documentation process.

References

- Alberola-Colomar, Pilar (2004): "La macroestructura del folleto bancario como instrumento de persuasión: estudio contrastivo inglés-español". *Ibérica* 8 (2004): 41-63.
- Alcaraz Varó, Enrique and Brian Hughes (2003): *Diccionario de Términos Jurídicos. Inglés-Español, Spanish-English*. Barcelona: Ariel.
- AVUEX Blended course (2006): "AVUEX: Proyectos fin de carrera-Informática". Cáceres: Polytechnic school, University of Extremadura.
[<http://campusvirtual.unex.es>].
- Biber, Douglas (2006): *University Language. A Corpus-based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.
- Biber, Douglas, Susan Conrad and Randi Reppen (1998): *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Borja Albi, Anabel, Silvia Gamero Pérez and Juan Carlos Ruiz Antón (1999): "The Translation Laboratory as a Learning Centre". *Translation. Nouvelles de la F.I.T.* 4 (1999): 439 - 453.
- Bowker, Lynne and Jennifer Pearson (2002): *Working with Specialized Language. A Practical Guide to Using Corpora*. London: Routledge.
- Chapelle, Carol A. (2001) : *Computer Applications in Second Language Acquisition*. Cambridge: Cambridge University Press.
- Collin, S.M.H., Jordi Piqué, Santiago Posteguillo and Lourdes Melción (2004): *Diccionario Bilingüe de Informática Inglés - Español / Español - Inglés. Spanish Computing Dictionary English-Spanish / Spanish-English*. London: Bloomsbury.
- Cruz Cabanillas, Isabel de la, Cristina Tejedor Martínez, Mercedes Díez Prados, and Esperanza Cerdá Redondo (2007): "English loanwords in Spanish computer language". *English for Specific Purposes* 26 (1): 52-78.
- Curado Fuentes, Alejandro (2000): *A Lexical Common Core in English for Information Science and Technology*. Cáceres: Servicio de Publicaciones, Universidad de Extremadura.
- Errico, Elena and Mara Morelli (2005): "Las lenguas para fines específicos en la formación de los mediadores lingüísticos entre español e italiano en Italia". In A. Curado et al., eds., *Languages for Academic and Professional Purposes in the 21st Century University*

- Framework. Cáceres: Servicio de Publicaciones Universidad de Extremadura, 451-458.
- Falcón, Pompeya (2000): "Traducción de artículos de investigación en Química desde una perspectiva del análisis del discurso". In F. Luttikhuisen, ed., *Actes III Congrès Internacional sobre Llengües per a Finalitats Específiques*. Barcelona: Universitat de Barcelona, 125-130.
- Francis, Gill (1986): *Anaphoric Nouns*. Birmingham: University of Birmingham.
- Fuertes-Olivera, Pedro A. (2005): "Traditional methods in specialised lexicography". *Ibérica* 9 (2005): 149-159.
- _____. (2007): "Metaphor in specialised discourse: An analysis of farming verbs in the WBE corpus and some Business English dictionaries". In A. Curado Fuentes, P. Edwards Rokowski and M. Rico García, eds., *Approaches to Specialised Discourse in Higher Education and Professional Contexts*. Newcastle: Cambridge Scholars Publishing, 2-20.
- Hoey, Michael (2005): *Lexical Priming. A New Theory of Words and Language*. London: Routledge.
- Leitchik, Vladimir and Sergey Shelov (2003): "Terminology: Where is Russian science today?". *LSP & Professional Communication* 3 (1): 82-110.
- López García, Ángel (2005): "Conexiones informáticas y conexiones neuronales: El problema de la traducción". In M.L. Carrió Pastor, ed., *Perspectivas Interdisciplinarias de la Lingüística Aplicada*. Valencia: Universidad Politécnica de Valencia, 31-42.
- Lozano Palacios, Antonio (2002): *Vocabulario Inglés/Español-Español/Inglés para los Estudios de Biblio-Documentación*. Granada: Universidad de Granada.
- Luzón, M^a José (2005): "El concepto de "colonia de géneros" en la enseñanza de Inglés para Fines Específicos". *Ibérica* 10 (2005): 133-144.
- McCarthy, Michael (1998): *Spoken Language & Applied Linguistics*. Cambridge: Cambridge University Press.
- McNaught, John (1993): "User needs for textual corpora in natural language processing". *Literary and Linguistic Computing* 8 (4): 227-234.
- Ooi, Vincent B.Y. (1998): *Computer Corpus Lexicography*. Edinburgh: Edinburgh University Press.
- Osborne, John (2004): "Top-down and bottom-up approaches to corpora in language teaching". In U. Connor and T. A. Upton, eds., *Applied Corpus Linguistics. A Multidimensional Perspective*. Amsterdam: John Benjamins, 251 - 266.
- Pisanski-Peterlin, Agnes (2005): "Text-organising metatext in research articles: an English-Slovene contrastive analysis". *English for Specific Purposes* 24 (3): 307-319.
- Posteguillo-Gómez, Santiago (2003): *Netlinguistics. Language, Discourse and Ideology in Internet*. Castelló: Servicio de Publicacions, Universitat Jaume I.
- _____. (2005): "Editorial". *Ibérica* 9 (2005): 5-6.
- Sankoff, David, Shana Poplack and Swathi Vanniarajan (1990): "The case of nonce loan in Tamil". *Language Variation and Change* 2 (1990): 71-101.
- Scott, Mike (1997): "PC analysis of key words and key key words". *System* 25 (1): 1-13.
- _____. (2000): *WordSmith Tools 3.0*. Oxford: Oxford University Press.
- Sinclair, John (2004): *Trust the text: Language, Corpus and Discourse*. London: Routledge.
- Slaouti, Diane (2002): "The world wide web for academic purposes: Old study skills for new?". *English for Specific Purposes* 21 (2): 105-125.
- Tsiplakou, Stavroula, Yanna-Maria Panavi and Yolanda Pandeli (2004): "Code-switching in computer-mediated communication and 'electronic performativity': the case of Greek email users". In S. Bolton, ed., *Proceedings of the 9th International Pragmatics Conference*.

Vancouver: Public Knowledge Project & John Benjamins, 1-15.

<http://webhost.ua.ac.be/tisp/papers.php>

Verdejo-Segura, M^a Mar (2003): “El uso de la traducción en la enseñanza de inglés para turismo”.
In P. Durán et al., eds., *Las Lenguas para Fines Específicos y la Sociedad del Conocimiento*.
Madrid: Universidad Politécnica de Madrid, 337-345.

Yus, Francisco (2005): “The linguistic-cognitive essence of virtual community”. *Ibérica* 9 (2005):
79-102.