

Towards a Lexical Framework for CLIL

John Eldridge, Eastern Mediterranean University (Turkey)

Steve Neufeld, Middle East Technical University (Turkey)

Nilgun Hancioğlu, Eastern Mediterranean University (Turkey)

Abstract

Research into lexical patterning and frequency has grown apace in recent years, largely due to developments in computer software that have enabled the ready processing of extremely large banks of linguistic data. The insights that are emerging from this research have profound implications for content and language integrated learning. In this paper, we explore how a principled and corpus-informed approach to vocabulary learning can be integrated into the CLIL framework. In particular, we attempt to find answers to the perennial questions of what types of vocabulary CLIL practitioners should be teaching, in what order they should introduce it, and what kind of approaches to the teaching and learning of vocabulary might best lay the groundwork for success in the CLIL environment.

Keywords: vocabulary, lexical syllabus, corpus-informed approach, word frequency lists, Common European Framework of Reference

1. Introduction

CLIL practitioners have long since noted the dilemma that emerges when learners seem to have insufficient linguistic resources to cope with the content in hand. As Clegg (n.d.) remarks, the choice seems to rest between adjusting the content of delivery, or the language. In the case of the former, the dilution of content learning objectives would seem to immediately raise doubts about the efficacy of the CLIL approach. In the case of the latter, it would seem essential to consider questions of how to grade and sequence language learning within the CLIL environment, so that the language integrated element of the programme not only provides short term coping strategies but lays the foundation for long term success in the academic environment.

In this regard, practice in CLIL has not perhaps informed itself as fully as it might have into ongoing corpus-informed research into the English lexicon, research that provides profound insights and serious implications for all CLIL practitioners.

2. Word Frequency Lists

It is commonly estimated that an educated native speaker of English has a receptive knowledge of between around 17,000 and 20,000 word families (Cervatiuc, 2008). Of these, Nation and Waring (2004) suggested that a receptive knowledge of the most frequent 6000 word families is sufficient to provide a working understanding of the language. This is on the basis that to understand a given text in English without assistance, and to be able consistently to infer the meaning of unknown words requires prior knowledge of around 95% of the items in the text.

The compilation of extensive databanks, or corporuses of language in use, such as the British National Corpus (2005), has enabled the production of wordlists that identify not only what these word families are, but their frequency of use, and further, how exactly they are used and in what particular contexts. Some interesting statistics emerge from the process. As reported by Cobb (n.d.), the most frequently occurring 1000 words of English will on average account for 72% of any given text. The most frequently used 2000 words of any given text will account for 79.7%. The most frequently used 3000 words of any given text will account for 84%, and the most frequently used 6000 words will account for 89.9%. As impressive as this sounds, this is still below the threshold of 95% suggested for unassisted reading.

Systematically learning the most frequent words in English as early as possible in the educational process would thus seem to be one of the most essential goals of any type of language instruction, CLIL included. However, at the same time, it is worth emphasizing that as learners work their way down these frequency bands there is a progressive and substantial decrease in learning gain. The effort put into learning the first 1000 words yields a very high return on the investment of time and energy. The second 1000 words however provide a much lesser return, and each successive band incrementally less again.

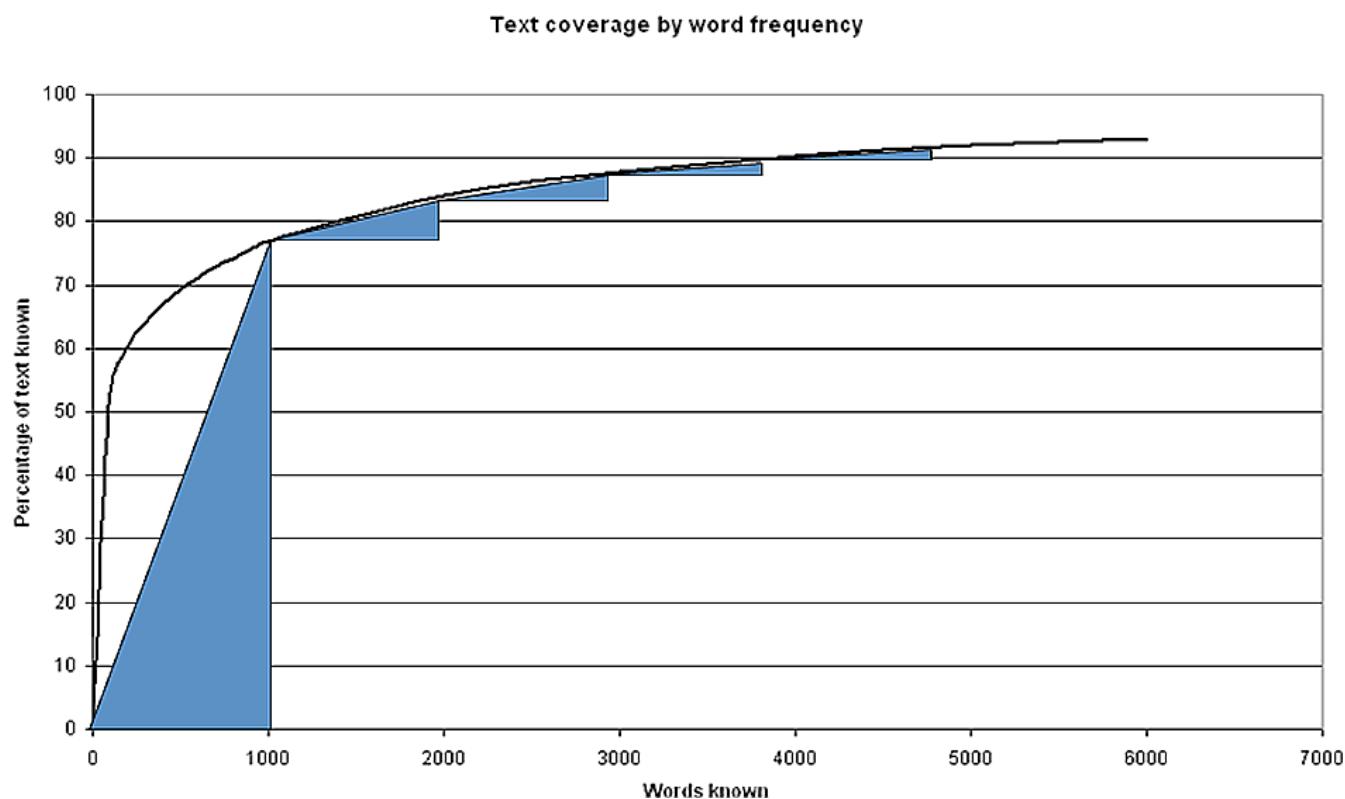


Figure 1: The law of diminishing return in learning vocabulary beyond the most common words of English

Given also that it has been further estimated that learners need 10-12 exposures to a word before they start to remember it (Nation, 2001), and given also that as a learner proceeds down the frequency lists, words start to occur less and less frequently, we can see that there is a plausible explanation both for the rather fast initial acquisition of a second language and the painstaking lack of progress that often seems to stifle and frustrate language learners and teachers from the post-elementary level: the lower frequency but still essential lexis is simply not occurring with the frequency that is enabling learners to absorb it with the facility that they acquired the very high frequency vocabulary.

Attempts to provide a high economy solution to this problem have included the identification of a 570 word family strong academic word list (Coxhead, 2000) that was compiled as an add on to the most frequent 2000 words of English as delineated in the General Service List (West, 1953). Initial results suggested that the combination of the two lists should enable familiarity with around 90% of the word families in academic text (Cobb). Hence the Academic Word List should be of considerable interest to CLIL practitioners. Subsequent research however, for example, by Hancioğlu et al (2008) strongly suggests that many of these so-called academic words are in fact relatively frequent in general English, and that the 2000 cut-off point taken by Coxhead for 'general' English was in fact too low a point for any genuinely authentic academic word list to be compiled. This is not to say the words on the AWL are not useful words. Indeed they are, but the underlying implication that a 2570 word vocabulary would enable a learner to cope in an academic environment seems unlikely to be viable in practice.

3. Lexis and CLIL

There is perhaps a natural tendency in a CLIL environment to start from the assumption that key lexis is basically defined by the content of the subject in question (Feldman & Kinsella, 2005). However, considering the research we have summarized, successful and thorough implementation of CLIL almost certainly requires by its end point:

- i) Knowledge of around the 6000 most frequent words in English.
- ii) Knowledge of the key lexicon of the content area.
- iii) Knowledge of the key transactional lexis of the educational environment, including knowledge of the key lexis used by digital media.

It should be noted that these are not discrete and distinct categories as such, but categories with a high degree of lexical overlap. At the same time, an imbalance in lexical focus can quite clearly have the effect of leading to lexical deprivation in one or more of these areas and hence directly to learning and teaching difficulties.

On the more positive side, the fast-mapping model of McMurray (2007) suggests a distinct lexical threshold of around 1600-1700 of the most frequent word families. Students who have naturally acquired these words in the course of their language learning generally seem to have acquired an actual vocabulary size of around 6000 words. However, students who fall even 200 or 300 word families below the threshold seem to have a vastly reduced vocabulary in total and consequently find it extremely difficult to cope with content studies in the medium of English.

Whilst similar thresholds may be assumed to exist in terms of content and transactional language, it seems evident that CLIL must take account, and early account, of the need for students to acquire as quickly as possible a firm knowledge of the most frequent words in the English lexicon, as well as basic content and transactional lexis.

4. Learning the Most Frequent Words

As already noted, the most frequent lexis in English occurs with such frequency that it would appear to be acquired with relative ease. Very quickly however, vocabulary development starts to become increasingly stressful, almost certainly because the frequency of occurrence of new vocabulary is continually decreasing. Although it might be thought that language teaching pedagogy would have firmly addressed this problem, studies of contemporary English language teaching course books seem to repeatedly show that the explicit focus on vocabulary throughout entire series of course books from beginner to upper-intermediate consistently falls beneath the threshold suggested by the fast-mapping model. (Cobb, 1995; Eldridge & Neufeld, 2009). Furthermore, despite purporting to adopt a lexical approach, EFL course books on the whole not only clearly fail to provide coverage of the most frequent words in English, but also fail to provide sufficient systematic recycling and repetitions of key words to facilitate long-term acquisition.

A popular solution to this, the encouragement of extensive natural reading for vocabulary development is beset by a similar shortcoming. In several studies of series of graded readers, (Cobb, 2007; Eldridge & Neufeld, 2009) it was again identified that key high frequency words did not occur with sufficient regularity to facilitate natural acquisition. This does not mean that extensive reading is without value, merely that it is not necessarily the most efficient way of promoting vocabulary acquisition. A systematic approach to the teaching and learning of lexis in a CLIL environment must then not only identify a key vocabulary but ensure that exposure to that vocabulary and opportunities to use that vocabulary are maximized to levels that will genuinely contribute to ease of learning.

It is worth remembering also that the rather vague notion of what it means to actually 'know' a word also needs some further extrapolation. At the most basic level, it may mean little more than recognizing and understanding a word when we see it. Gardner (2007) explores this issue in depth, but suffice it to say that at a more advanced level, it means knowing the members of the word family, knowing the lexico-grammar of the word, other words with which it occurs (collocates) and more. The productive knowledge that enables fluent, appropriate and accurate use of a word requires a level of attention and intensity of focus way beyond that required to work one's way through a text, and broadly comprehend its subject matter. Simple provision even of the 10-12 exposures discussed earlier may help with the recognition of a word, but it is unlikely to be sufficient to facilitate fluent use

in writing or speaking. And whilst a wide receptive vocabulary may be a very reasonable objective in some areas of language learning, it is hardly a very tenable objective for the highly communicative environments that CLIL tends to inhabit.

5. Identifying Key Vocabulary in CLIL

As part of their ongoing research into lexical frequency, Billuroğlu and Neufeld (2005) took a number of common word frequency lists and by identifying the commonalities between them, produced a new list, comprising the 2709 most frequent word families in English. This list, and indeed other similar lists, if used systematically, can certainly help provide the general foundation that will take learners way beyond the fast-mapping threshold identified by McMurray (2007).

If CLIL practitioners then adopt a simple corpus-informed approach to their subject matter, the second part of the puzzle should quickly be elucidated. By feeding content matter such as core books, texts and notes into what are now freely available concordancing software and word frequency calculators, a content-based wordlist will quickly be generated that identifies the key content vocabulary. What this process will reveal inevitably is that the content list will contain a marked overlap with general frequency lists such as the BNL, whilst at the same time identifying the key conceptual terms that form the bedrock of the subject matter. The statistics should clearly show also that for learners to succeed in their studies – and the lower their language level the more critical this will be – they must be familiarized with the core vocabulary of general English that constitutes the scaffolding for all communicative activity in the language, regardless of subject matter.

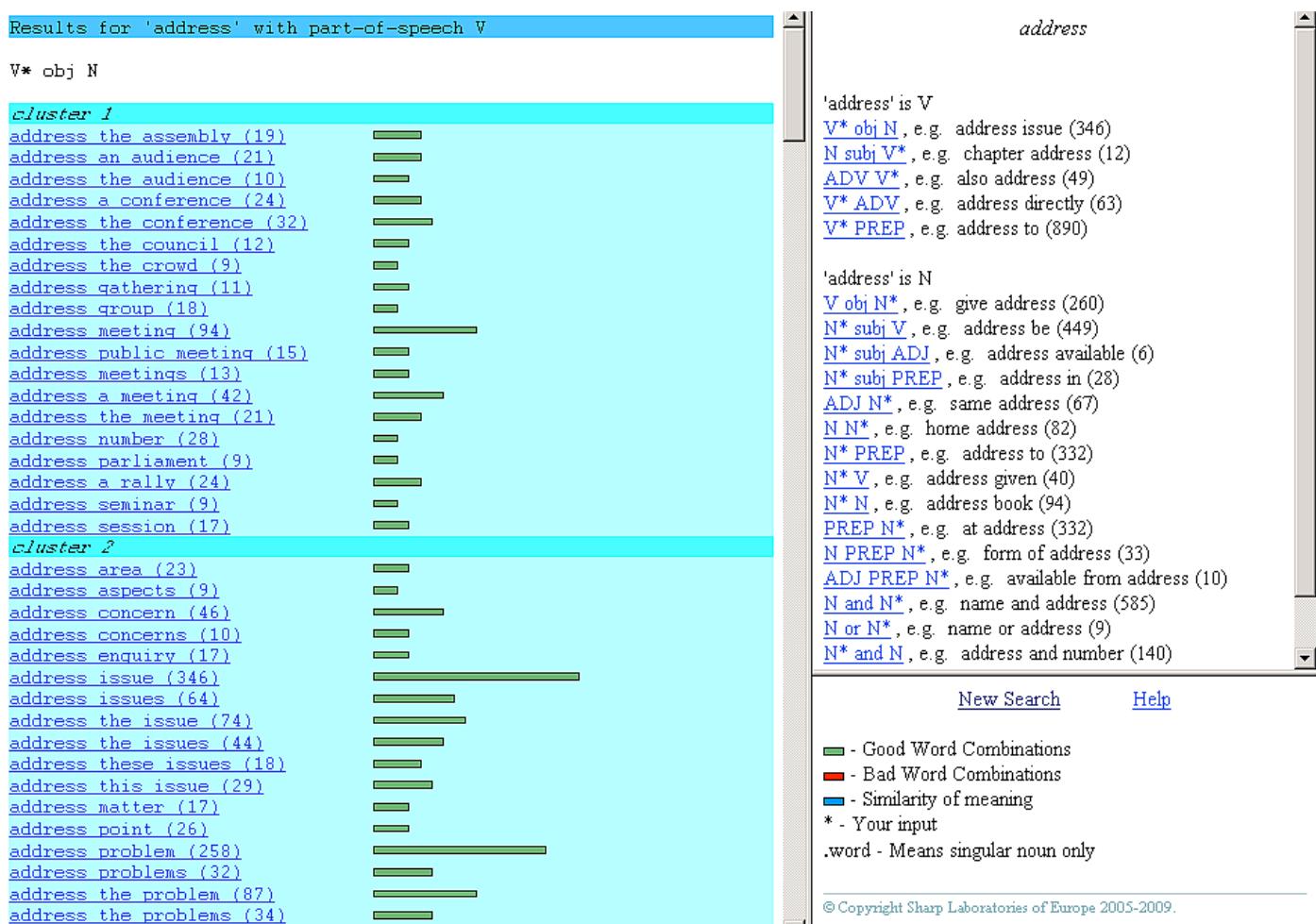


Figure 2: Just the Word (Sharp Laboratories of Europe, 2009) output for the word 'address' showing the 'depth' and 'breadth' of lexical knowledge involved in 'knowing' a common word.

At the same time, concordancing software will also identify subject-specific uses of words that occur in general English. The word *table* is a frequent enough word in general English, but a *table* in mathematics means something else rather obviously, as does *tabling a motion*, or *water table*. A content-based lexical framework thus needs to account for both the general and the specific and to be highly alert to the fact that content and language integrated learning cannot define activity around the relatively narrow set of lexical items that seem to somehow intuitively form the basis of the subject matter. The transactional language of the classroom, the instructional discourse of, say, examinations, and the interface language of computers all present similar cases, and are open to similar treatment.

6. Sequencing Vocabulary Teaching in CLIL

Many language teaching syllabuses work from a pre-formulated syllabus staged around grammatical structure, communicative function and/or language skills. Any systematic approach to vocabulary development is forsaken for the usual lists of 'new' words and phrases that emerge helter skelter from the eclectic readings often chosen to attract more than instruct. The recent trend in publishing has been to map course books into the Council of Europe (2009) Common European Framework of Reference for Languages (CEFR), which has provided detailed descriptors for six levels of language proficiency across all four major language macro-skills and in three different dimensions of operation: work, study, and social. The CEFR is a description of language proficiency that can apply to any of the major European languages, as exemplified by the Association of Language Testers in Europe. 'can do' statements (2002). As the CEFR is designed to be general enough to describe language abilities in any foreign language, the issue of lexis and vocabulary is naturally omitted, as the lexical structure of each language is unique. So, while the CEFR is providing a great tool to standardize approaches to teaching and assessing any foreign language from a skills basis, the lack of lexical frameworks for individual languages has driven vocabulary development further and further to the hinterland of language learning and teaching.

Some recent developments in course design do attempt to give vocabulary equal weighting with other course elements. The best example is Nation (2007) who advocates the importance of high frequency words to support an equal balance between the four strands of meaning focused input, meaning focused output, focus on form and focus on fluency. The high frequency words in this case are based on the written word, and determined by their use in receptive and not productive knowledge of the language. Bauer and Nation (1993) meanwhile postulated seven levels of word formation to define the nature of a word 'family' for research into receptive vocabulary knowledge. According to the strict application of Bauer and Nation levels, having learned the root word BOOK, a learner would then recognize the forms BOOKS, BOOKED, and BOOKING in a reading passage. Whether they would actually know the difference in meaning between "the police booked the suspect for fraud", "we booked our holiday early", and "the book value of a second hand car", is another matter, depending on their knowledge of the words providing the context (e.g. police, suspect and fraud in the first example.)

However, from the practitioner's point of view, it would be highly useful if at least some kind of lexical framework for the target language could be provided to coincide with the levels of language ability as defined by CEFR. To fit with the CEFR, which is described by what students 'can do', such a lexical syllabus would take the shape of word lists that describe the productive ability, i.e., what students 'can say' and 'can write', not merely what they can recognize when they see it. In the following section we will describe a preliminary attempt to create such a list.

7. Developing the Common English Lexical Framework (CELF)

In designing The Common English Lexical Framework (CELF), the prime objective was to produce a lexical syllabus describing the words that a learner should be able to use productively according to their language proficiency as described by the Common European Framework of Reference for Languages (CEFR).

In order to establish this broad framework, the following procedure was followed:

- i. The Rinsland (1945) corpus of 6 million words of students' written work collected across grades 1 to 8 was used as the basis to determine the productive range of lexis of children in the formative years of schooling in America. Although the corpus is dated, word frequencies tend not change very much over time. The data West used to generate the GSL in 1953 came from corpora compiled in the 1920s and 1930s, yet the GSL is still in use today. The foundation of modern readability statistics (Dale & Chall, 1948) is largely based on linguistic data from the 1940s. The Rinsland corpus therefore provided an excellent framework for further analysis.
- ii. The words that emerged as keywords in the corpus were then further investigated and classified by comparing them with the frequency of use of words in other extant lists. These lists included:
 - a. 2,500 most common words spoken by children (Stemach & Williams),
 - b. the 3,000 most common words from grades 1 to 5 in America (Dale & Chall),
 - c. the 220 most common 'sight' words (Dolch sight words),
 - d. the 20,000 most commonly used words in British English (Kilgarriff) and
 - e. the 2,709 most commonly used words in English in the Billuroğlu and Neufeld List (2007).

Investigation of word frequency in general, word frequency in specific age groups and at particular levels, and an examination of the CEFR 'can-do' statements thus enabled a categorization of the resulting word families into the six discrete levels of the Common European Framework.

One of the features of CELF that distinguishes it from other word lists is the staging of members of a word family. CELF maintains the concept of a word family as in BNL2709 (Billuroğlu & Neufeld, 2007), unlike pure frequency lists (Kilgarriff). However, rather than assigning the same ranking to all word family members of one family, CELF ranks each word family member according to its active use in productive writing that corresponds to the CEFR bands. So, although we started from the premise that words are listed according to word families (e.g. FACE is the 'headword' of a family made up of four words: face, faces, faced, facing), the individual words themselves need not appear in the same CEFR ability-linked band. The word family member FACE, for example, is used actively at CEFR-A1 but another word family member FACED requires different lexico-grammatical contexts and only appears in productive use at CEFR-B1. This depth of analysis is what gives CELF its status as a lexical syllabus, as opposed to a list of words.

The word family breakdown according to CEFR levels is shown in the table and chart below. As implied earlier, the newly produced CELF can be the basis for a lexical syllabus that can be applied alongside a CLIL approach in keeping with the Common European Framework in terms of assessed language ability. In order to be functionally bilingual, students should probably be able to function at a B2 level in terms of the CEFR skills descriptors and at the same time have a good productive knowledge of the most common meanings and uses of the words from the corresponding CELF lexical syllabus.

The CELF lexical syllabus does not suggest that these are the only words students need to know. Each student will also have their own individual lexicon consisting of words they have found useful in their own context, as well as the 'special' words in their subject areas.

	Headwords	Family Words
A1	692	1154
A2	567	1376
B1	604	1630
B2	526	1847
C1	393	2130
C2	434	1825

Table 1: Headwords and family words that students should be able to use productively at CEFR bands of language ability

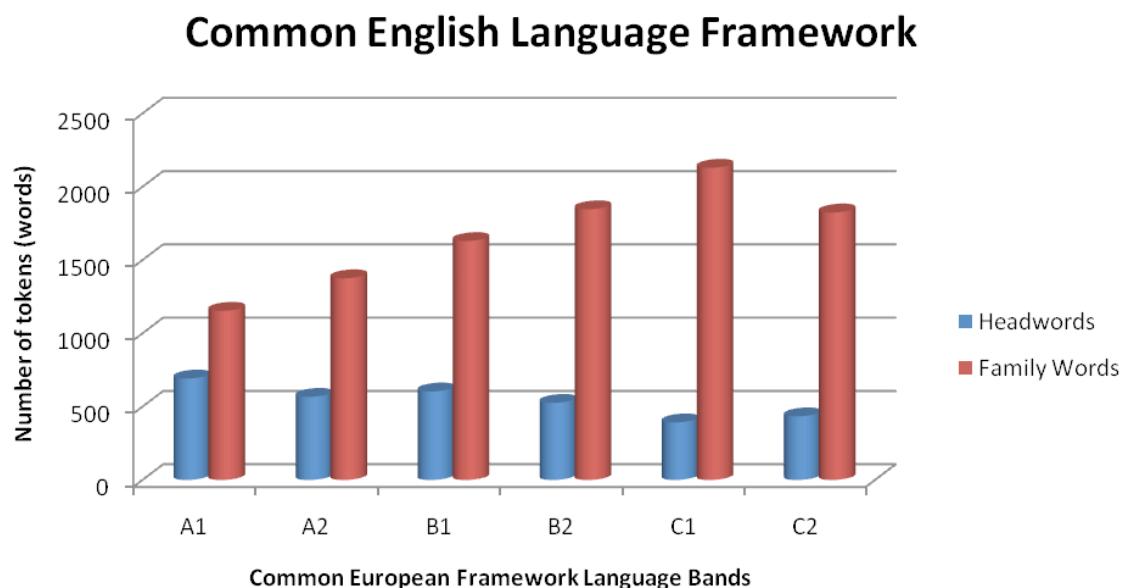


Figure 3: Headwords and family words that students should be able to use productively at CEFR bands of language ability

Many of the words in the CELF are 'multi-purpose' words, and need to be explored in both depth and breadth. In particular, words that may appear to be general words may also have very specific meanings in the context of specific subjects. For example, 'SCALE' has distinct meanings, uses and collocations in Mathematics, Geography, Geology, Music, and Biology. Indeed on further analysis and investigation, it transpired that there are actually 3,605 words in the CELF that have specific meanings within the broad subject areas of study in primary and secondary school as categorized by the MacMillan School Dictionary (Rundell, 2004). Thus in the next stage of the project, the words in the CELF having already been divided into six levels were further tagged according to possible applications in content study. The key below indicates the range and extent of this operation:

Mus	Music
Edu	Education
Sci	Science
Phy	Physics
Lan	Language
Hea	Health
Mus	Music
Ana	Anatomy
Env	Environment
Soc	Social Science
Bio	Biology
Mat	Maths
Com	Computing
Agr	Agriculture
Che	Chemistry
Rel	Religion
Eco	Economics
Lit	Literature
Art	Art
Geo	Geography/Geology
Ast	Astronomy
Soc	Social Science

Table 2: Key to subject areas for CLIL

The principled approach in the undertaking described above has thus generated a lexical syllabus that provides target vocabulary in English for each of the six levels in the Common European Framework, and provides guidelines for curricular sequencing and implementation. By further identifying general English words that have specific subject meanings, a further bank of support has been provided that will enable CLIL practitioners to explore the shifting nature of lexis across the curriculum as they work with these words. What remains is to consider how the learning of the words can best be facilitated in the CLIL environment.

BNL	HeadWord	A1	A2	B1	B2	C1	C2	CLIL
1	off	off						Mus
1	office	office	officer	officers, offices				
1	often	often						
1	oil	oil	oiled		oils, oily		oiling	Art, Bio, Che, Env
1	old	old	older		oldest		olden	Art, Rel
1	once	once						
1	one	one, ones		one's		oneself		
1	only	only						Com
1	open	open, opened	opening, opens		openings	opener		Eco, Edu
1	other	other, others						
1	own	own	owned, owns	owner	owners	ownership		
1	paper	paper, papers				papered, papering		Che, Edu, Mat, Sci
1	part	part	parts	apart, particles	parted, partial, parting, partly	partially, particle		Lan, Phy, Sci
1	party	party		parties				
1	pass	pass	passed	passing	passes			Edu, Geo
1	past	past						Lan
1	pay	paid, pay	paying, pays	payment, repay	repaid	payments		Edu
1	perhaps	perhaps						
1	person	people	person	personal, persons	peoples, personality, personally			Com, Lan, Soc
1	picture	picture, pictures						
1	piece	piece, pieces		apiece				
1	place	place	placed, places	placing		displace		Mat, Phy
1	plant	plant, planted, plants	planting	planter, planters				Agr, Bio
1	play	play, played, playing, plays	player, playful	players, playground, playgrounds				Edu, Lit, Mus
1	please	please	pleasant, pleased	pleasing, pleasure, pleasures, unpleasant	pleasantly, pleases	displeased		

Figure 4: Extract from the Common English Lexical Framework, printed version.

8. Towards a Lexical Approach to CLIL

A fundamental thesis of this paper is that lexis is the foundation of language, and a fundamental determinant of how well CLIL works in practice. We have suggested that the research base for identifying key vocabulary for CLIL is fundamentally intact, and reported on research suggesting how this vocabulary can be sequenced in the teaching-learning process. We now need to look at more depth into the issue of how learners can be provided with sufficient exposure to and practice of that vocabulary to ensure long-term acquisition, especially given the rather chilling possibility that 10 to 12 formal encounters with a word may be sufficient only to facilitate receptive understanding. The following 'LexiCLIL' principles provide some food for thought and exploration:

- i. Key to success in a CLIL environment is the acquisition of a productive vocabulary that includes knowledge of
 - the most frequent vocabulary items in the target language.
 - key vocabulary in individual subject areas.
 - key vocabulary needed to function in the educational environment.
- ii. A coherent and economic approach to vocabulary acquisition requires a coordinated and systematic approach that functions across the curriculum.
- iii. The bands of the Common European Framework for languages and word frequency lists such as the BNL and CELF provide a firm basis for the staged acquisition of vocabulary to be built into the curriculum.
- iv. All lessons present opportunities for vocabulary learning, recycling and production opportunities
- v. Vocabulary cannot just be 'picked up'. Repeated exposure and practice of key words is vital.
- vi. Vocabulary almost certainly needs to be an integral part of assessment in all subjects. The question of a CLIL approach to assessment needs to be examined.
- vii. The Internet and Web 2.0 tools offer unparalleled opportunities to enrich vocabulary teaching and learning and should be embedded in a LEXICLIL approach.

Having outlined these broad principles for consideration, in the final sections of this paper, we will briefly turn our attention to two specific methods that practitioners might find of benefit in developing strategies for the acquisition of key lexis.

9. MOODLE and Electronic Readers

The increasingly popular MOODLE learning platform has been explicitly built on a social constructivist philosophy, and as with many Web 2.0 tools allows users considerable freedom to interact and create text for themselves.

Topic outline

Welcome to the Great *Animals* eReader!

This E-Reader introduces you to some of the most extraordinary animals on our planet. Unfortunately, like the Giant Panda, all of them are in danger. Let's learn more about these animals and what we can do to protect them.

Before we start though, take some time to introduce yourself to some of the other members of the Great Animals E-Reader community.

We recommend that you use Internet Explorer with eReader. Some activities in the eReader will not work properly in other browsers like FireFox or Chrome.

[Orientation to the E-Reader](#)

[Key BNL2709 words](#)

Meeting point and Discussions

As you work through the eReader, feel free to participate in any of the discussions below. These discussions will provide you with further practise and help you develop your English as well as giving you the opportunity to get in touch with other learners working in this eReader.

- [News Forum](#)
- [Introducing Yourself](#)
- [Remarkable Animals Discussion](#)
- [Animals in our countries Discussion](#)
- [Saving our Wildlife Discussion](#)
- [What can we do to protect our animals Discussion](#)

1 The Cheetah

Research before you read

- [Word cloud about Cheetahs](#)
- [Watch a video about the cheetah](#)
- [Pre-Reading Activities](#)

Reading and Learning



LexiTronics

Nominated

ELTONS 09 CELEBRATING INNOVATION

BRITISH COUNCIL

CAMBRIDGE DICTIONARIES ONLINE

Blog Menu

- Add a new entry
- View my entries
- Blog preferences
- View course entries
- View site entries

Discovery News - Animals

Discovery

Cheetah - Social Function - Socially Active

Cheetahs are built with a light frame, global warming and ice age **habitat**, according to scientists.

Animals that have the best **life** success!

Animals that spread information using a preferred side of the brain **love** humans and humanizing.

Go! Go! Go! Play! Play! Walk! Walk! Animals

Animals have a hard world **survive** may benefit from your guidance.

Birds Fly! Fly! Walk! Walk! Wings

Flamingos have blue feathers because their wings were exposed so they can't **protect** their wings.

Seal! Seal! Seal! Seal! I like my food! >

Seals exposed to oil pollution are **forced** to have smaller testicles than seals.

Figure 5: CLIL in the guise of an eReader about Great Animals in MOODLE

Using MOODLE we created a series of texts about animals, as shown in Figure 5 above, by adapting the texts from Wikipedia, and then isolating a core set of key vocabulary from the BNL 2709 (Billuroğlu & Neufeld, 2007) for intensive focus. During the adaptation phase, it was then ensured that the key target items all occurred at least twelve times through the series of five texts. Using some of the tools offered by the MOODLE platform, we were then able to create an interactive reader that functioned as follows:

1. Learners are encouraged to prepare for the reading by conducting research into the content in their own language, thus building up their content knowledge in preparation for revisiting it in the second language.
2. A glossary entry was created for each target item. Within the glossary, automated links were provided to online dictionaries, and activities related to each target word set up to function in Blogs and WIKIs.
3. The MOODLE glossary feature highlights the target words wherever they appear on the course page, and links the students to the glossary, where they find further information about the word and activities to do.
4. As the students do the activities in the public space of the blogs, wikis and discussion forums they become creators of text for other students to read. The glossary tool continues to automatically highlight the key words, and thus within the dynamic medium of the learning platform, instances of the key words start to multiply further, as the learners essentially create learning opportunities for each other.
5. Further practice and rehearsal is provided through simple automated text reconstruction software.

This is a simple example of how the dynamism of Web 2.0 can enable us to maximize opportunities for vocabulary acquisition, and even more importantly enable learners to create those learning opportunities for themselves.

10. ANKI Flashcards

Interactive flashcards can function in a similar way. Using the Common English Lexical Framework, packs of cards were generated for each of the lexical items that had been identified from the research, with each band of lexis being divided into sets of 200 cards according to the level of difficulty of the word. The flashcard decks are viewed in the ANKI (Elmes, 2009) flashcard system, which is an open source software that uses a 'spaced repetition system' that presents words for review according to each student's need to review and recall.

The decks can be downloaded from Lexitronics at <http://lexitronics.edublogs.org>

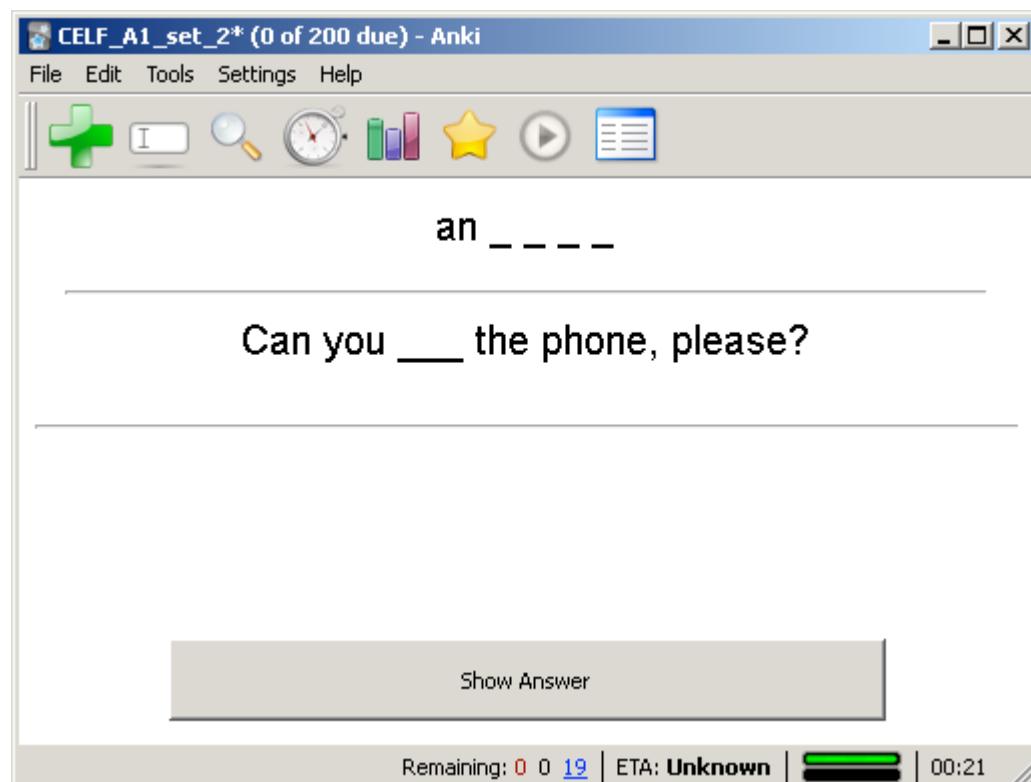


Figure 6: 'Front' of ANKI flash card, showing the clues for the target word.

As noted before such cards may be particularly useful in early stage language learning in driving students towards the key thresholds they must pass if their language is to progress, providing contextualized and continued exposure to key lexis on a continued and repeated basis.

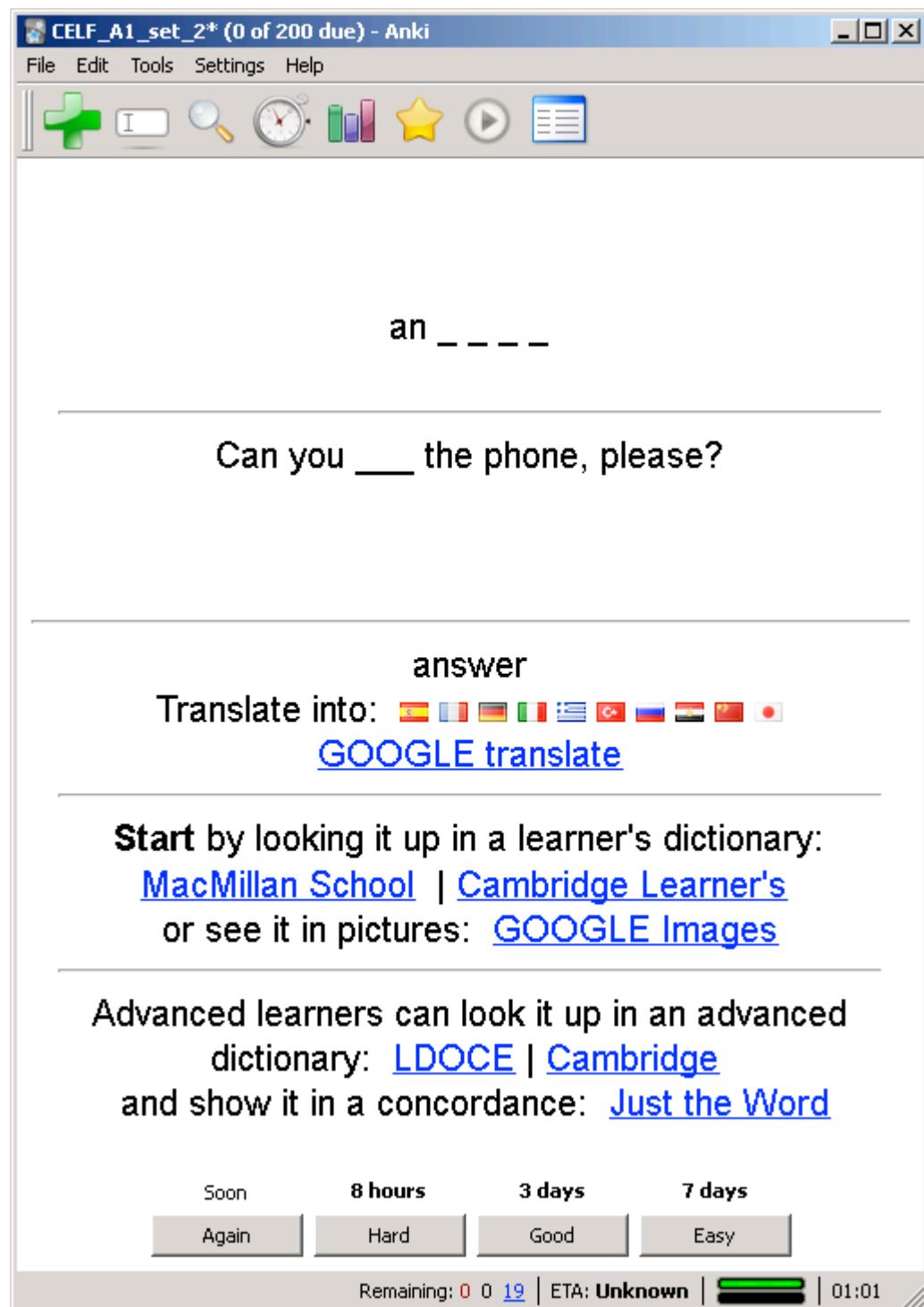


Figure 7: 'Back' of ANKI flash card revealed, with dynamic links to various resources for the learner to explore according to their level and knowledge of the target word.

11. Conclusion

In this brief article we hope to have shown that it is quite feasible to compile and sequence a genuine CLIL lexicon. We hope to have shown also that there are extremely good reasons for doing so that are based on sound research evidence and statistical data about lexical patterning in language. The two tools we have described are merely exemplars of the many resources available to us. The main point to make is that by developing a structured lexical syllabus with an appropriate methodology, based on careful and frequent recycling of key vocabulary, we should be able to systematically address the vocabulary problems that many learners face. In suggesting a corpus-informed approach to course design, we have further attempted to demonstrate how learning, and particularly vocabulary learning can be driven by data rather than intuition. Furthermore, such data need not be the sole preserve of researchers. The tools for corpus-informed work are mostly freely available on the web, and can be put to use by any CLIL practitioner wishing to exert a higher degree of contextual sensitivity to that exhibited in this research. Finally, although not discussed here, in a CLIL context the principles of LexiCLIL naturally leads to data-driven learning whereby learners become self-directed researchers in the language learning process, identifying for themselves the language they need to take their studies further, and thereby taking ownership of the learning process.

References

- Association of Language Testers in Europe. (2002). *The ALTE can do project: articles and can do statements produced by the members of ALTE 1992-2002*. Retrieved May 10, 2009 from http://www.alte.org/cando/alte_cando.pdf
- Bauer, L., & Nation, I.S.P. (1993). Word families. *International Journal of Lexicography*, 6(3), 1-27.
- Billuroglu, A., & Neufeld, S. (2005). *The Bare Necessities in Lexis: A new perspective on vocabulary profiling*. Retrieved February 19, 2009 from http://www.lextutor.ca/vp/tr/BNL_Rationale.doc
- Billuroglu, A., & Neufeld, S. (2007). *BNL 2709: The essence of English* (4th. ed.). Nicosia: Rüstem Kitabevi.
- BNC Consortium. (2005). *The British National Corpus*. Available from <http://www.natcorp.ox.ac.uk>
- Cervatiuc, A. (2008.) *ESL Vocabulary Acquisition: Target and Approach*. *The Internet TESL Journal*, 14(1), Retrieved June 13, 2009 from <http://iteslj.org/Articles/Cervatiuc-VocabularyAcquisition.html>
- Clegg, J. (n.d.) Providing Language Support in CLIL. *FACT Journal* 6, Retrieved August 15, 2009 from <http://factbg.hit.bg/issue6/f6-clegg.pdf>
- Cobb, T. (1995). Imported tests: Analysing the task. Paper presented at TESOL (Arabia). Retrieved June 6, 2008 from <http://www.lextutor.ca/cv/Al-Ain.html>
- Cobb, T. (2007). Computing the vocabulary demands of L2 reading. *Language Learning & Technology*, 11(3), 38-64.
- Cobb, T. (2008). What the reading rate research does not show: Response to McQuillan & Krashen. *Language Learning & Technology*, 12 (1), 109-114.
- Cobb, T. (n.d.). The compleat lexical tutor for data-driven learning on the web [Computer software] <http://www.lextutor.ca/>
- Council of Europe. (2009). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238.
- Dale, E., & Chall, J.S. (1948) A formula for predicting readability. *Educational Research Bulletin*, 27, 37-54.

- Dale-Chall Word List [Data file]. Retrieved from <http://www.interventioncentral.org/htmdocs/tools/okapi/okapimanual/dalechalllist.php>
- Dolch sight words. [Data file]. Retrieved from <http://www.mrsperkins.com/dolch-words-all.html>
- Eldridge, J. & Neufeld, S. (2009) The Graded Reader is Dead, Long Live the Electronic Reader. *The Reading Matrix* 9(2), September 2009
- Elmes, D. (2009.) ANKI - spaced repetition system. [Software]. Available from <http://ichi2.net/anki/download/index.html>
- Feldman, K & Kinsella, K. (2005.) *Narrowing the Language Gap: The Case for Explicit Vocabulary Instruction*. New York, N.Y.: Scholastic Inc. Retrieved July 19, 2009 from http://teacher.scholastic.com/products/authors/pdfs/Narrowing_the_Gap.pdf
- Gardner, D. (2007). Validating the construct of word in applied corpus-based vocabulary research: A critical survey. *Applied Linguistics*, 28(2), 241-265.
- Hancioglu, N., Neufeld, S., & Eldridge, J. (2008). Through the looking glass and into the land of lexico-grammar. *English for Specific Purposes* 27/4, 459-479 doi:10.1016/j.esp.2008.08.001
- Kilgarriff, A. Lemmatized BNC frequency list [Data file]. Retrieved from http://www.kilgarriff.co.uk/BNC_lists/lemma.num
- McMurray, B. (2007). Defusing the Childhood Vocabulary Explosion. *Science* 317 (5838), 3 August 2007, p. 631
- Nation, P. (2007). The four strands. *Innovation in Language Learning and Teaching*, 1, 1-12.
- Nation, P. (2001). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, P., & Waring, R. (2004). *Vocabulary size, text coverage and word lists*. Retrieved February 19, 2008 from <http://www.wordhacker.com>
- Rinsland, H. (1945). *A basic vocabulary of elementary school children*. New York: Macmillan.
- Rundell, M. (2004). *Macmillan School Dictionary*. Macmillan Education.
- Sharp Laboratories of Europe. (2009.) Just the word. [Software]. Available from <http://193.133.140.102/JustTheWord/index.html>
- Stemach, G., & Williams, W. (1988). *Word express: The first 2,500 words of spoken English*. Academic Therapy Publications, Incorporated
- Stemach, G., & Williams, W. *The first 2,500 words of spoken English* [Data file]. Retrieved from http://www.lex tutor.ca/vp/kids/kid_list_display/
- West, M. (1953). *A General Service List of English words*. London: Longman, Green & Co.