A role for lexical bundles in the implementation of Content and Language Integrated Learning programmes in Colombian universities

SERGIO TORRES MARTÍNEZ

Introduction

The production of content-appropriate discourse in academic English has proved a difficult task for learners in different disciplines in Colombian universities. There is plenty of blame to go around: among other problems, educators frequently point to the belief that language courses (isolated from university subjects) automatically promote linguistic skills transferable to a specific subject-language, the informal approach to subjects demanding high-skill communicative proficiency, the abuse of L1 informal register in both L1 and L2 academic writing and speaking; the idiosyncratic academic culture dominated by a blinkered return-of-investment thinking, and the lack of academic rigour leading to distinct increments in existing knowledge. In such a scenario, it becomes apparent that a practical, unambiguous approach is to be taken in order to facilitate the implementation of a more seamless integration of language and content in Colombian EAP classrooms.

In this paper, I want to suggest that precisely such a synthesis can be brought about by giving a central role to the learning and teaching of lexical bundles within the pedagogic framework of Content and Language Integrated Learning (henceforth CLIL). As we will see, lexical bundles provide learners with a down-to-earth characterization of discourse based not only on frequency of occurrence, but also on their intrinsic learnability as well as the implied promise of 'native-like' proficiency, resulting in the development of 'growth mindsets' (Mercer and Ryan, 2010; Mercer, 2012). CLIL for its part may help dispel the commonplace notion that content language must be lecturer-led metalanguage.

The article begins by defining lexical bundles (henceforth LBs) as a subset of the more general concept of formulaic sequences. After that, a brief outline of research evidence concerning the role of LBs in the improvement of discourse quality and awareness is presented. An overview of the use of LBs in the English language classroom is then provided. The next section introduces the concept of the Content-language Continuum (CLC), before going on to consider CLIL in more detail. The role of LBs in the CLIL context and the instructional strategies to be applied are discussed in the next two sections. Finally, an example of a LB tool for classroom use is described.

Formulaic sequences and lexical bundles

Formulaic sequences can be defined as multi-word segments susceptible to being stored and retrieved



SERGIO TORRES MARTÍNEZ is a language instructor and a freelance multimedia designer. He has created a number of tools designed to bridge the gap between technology and language teaching in Colombia. Among his main interests are interaction design, CLIL curricula design in English and French, and the definition of best practice in the teaching of spoken grammar in English. Email: surtr_2000@yahoo.es

doi:10.1017/S0266078413000151

 © 2013 Cambridge University Press

https://doi.org/10.1017/S0266078413000151 Published online by Cambridge University Press

from the mental lexicon at the moment of production (oral or written). These chunks of language facilitate encoding-decoding processes, resulting in a more fluent production of discourse. Given their potential to increase production speed and fluency, formulaic sequences have been deemed to play a significant role in developing 'native-like' proficiency in L2 acquisition. The plausibility of this assertion is confirmed by the fact that formulaic sequences such as idioms and speech formulas (e.g. *I beg your pardon!*, *Thanks for having me!* etc.) are more likely to fit specific types of communicative situations at a higher encoding-decoding speed, showing 'a processing advantage over creatively generated language' for non-native as well as native speakers (Conklin & Schmitt, 2008: 72).

The more recent development of corpus-based research methodologies has led to the automatic identification of formulaic sequences by computers. This in turn has served to facilitate studies on the pedagogical implications of specific types of formulaic sequences based on their co-occurrence in a particular register or academic field. Among the most important work in both of these areas is that done by Douglas Biber, who refers to these automatically identified formulaic sequences as *lexical bundles* (LBs). Lexical bundles are repeated strings of words of variable length (3, 4, and 5 word sequences being the most commonly studied) which are observed to occur above a particular threshold level in a given corpus. The threshold level itself is in fact entirely arbitrary, but many studies follow Biber et al. (1999) in specifying that they must occur at least ten times per million words, and in at least five different texts in order to avoid confusion with idiosyncratic or repetitive uses.

Psycholinguistic research on LBs

There are two main strands of research on LBs. One of these strands is psycholinguistic in orientation, and centres on the question of whether and to what extent L1 and L2 speakers can consistently recognize lexical bundles as cohesive, holistically retrievable semantic units, thereby providing a solid path for generalizations concerning their psycholinguistic validity. A good example of this tradition of research is Nekrasova (2009). This study consisted of two experiments employing two different measures to assess L1 and L2 English speakers' knowledge of lexical bundles as holistic units. Since previous research had demonstrated a disparate use of lexical bundles by L1 and L2 speakers, Nekrasova's research question was logically focused on the range of these differences: 'Do English L1 and L2 speakers differ in their knowledge of lexical bundles that serve different discourse functions?' (ibid.: 655). The bundles analysed in the study were classified (following Biber) into 'discourse organizers' (i.e. bundles whose function is largely metadiscursive) and referential bundles (i.e. bundles that express the main propositional content of a text).

The results showed that there is a significant gap between proficient users (L1 and L2 higher proficient learners) and lower English learners (intermediate users) regarding their use of referential bundles. Although it was difficult to determine (as Nekrasova concedes) whether lower proficiency learners were unable to recognize certain bundles or simply chose equivalent options to reconstruct discourse context, this research does show that the decision as to whether to choose specific bundles or not seems to depend on the interplay between variables such as register distribution and discourse function.

Academic usage of LBs in EIL classroom

The other main strand of LB research is much more descriptive in orientation and focuses on observing disciplinary variation in the forms and functions of LBs and the implications of these differences for classroom teaching. An influential paper in this tradition is Cortes (2004), which identifies significant variation in LB usage across two disciplines – history and biology – and across two groups of writers – professional academics and students. Crucially, Cortes concludes by making a strong argument for the explicit teaching of lexical bundles in the EAP classroom:

> As lexical bundles are very frequent in published academic prose, it is necessary to encourage students to use these expressions. Instructors could introduce some of these lexical bundles (for example, those that have text organizing functions) together with some of those discourse markers or connectors, going beyond single words or two-word combinations to longer phrases, phrase fragments, or clause fragments. Native and non-native speakers of English would benefit from increased awareness of the functions performed by lexical bundles when used by published authors. As a result, students should have a better understanding of academic readings and be able to use these bundles in a more appropriate way in academic writing. (Cortes, 2004: 420–4)

https://doi.org/10.1017/S0266078413000151 Published online by Cambridge University Press

What this suggests is that LBs may provide students with accessible ways of understanding the kinds of texts that they must learn to read and write in terms of discourse cohesion, register appropriateness and context realization. In other words, LBs have the potential to unveil internal semantic connections between elements in a text (the referential function), while providing clues about how the text is organized and about the type of register the speaker is faced with (the discourse organizing function).

The content-language continuum

I define the content-language continuum (CLC) (see Figure 1) as the speakers' (NSs, NNSs) use of specific LBs in the academic context (research papers, dissertations, etc.) to consciously engage in a continuum of unambiguous meaning negotiation (by means of lexical practice focused on research-oriented, text-oriented and participant-oriented bundles). In other words, focusing on the discipline-driven meanings inherent in particular LB usage patterns, users can integrate LBs receptively and productively into the skill-development continuum, resulting in the reinforcement of content and language.

The CLC is therefore the place of encounter and reciprocal exchange of content (as a formally externalized realization of the scientific thought, its methods, rituals and subjectivity by means of books, papers, research reports, etc.), and the language, understood as speakers' discourse-driven realization of content both externally and internally. This model seeks to move away from the somewhat rigid perspective of language-content separatism, as well as the undue emphasis ascribed to 'difference' and 'otherness' imposed on the speaker, through a focus on language as a social phenomenon. At the same time, this approach requires an inclusive model of instruction drawing on content-language awareness, learner-centredness and a process-oriented strategy. This is where Content and Language Integrated Learning comes in.

Language-content awareness: the role of LBs in CLIL

The design of content-based language learning programmes in Colombia is challenging at many levels. Firstly, both learners and educators are loath to accept the importance of linguistic competence in their academic and professional lives. Secondly, content-based language instruction is not perceived as an academic asset due to the fact that only a few university subjects are taught through the medium of English. Finally, CLIL is considered a cost-prohibitive immersion

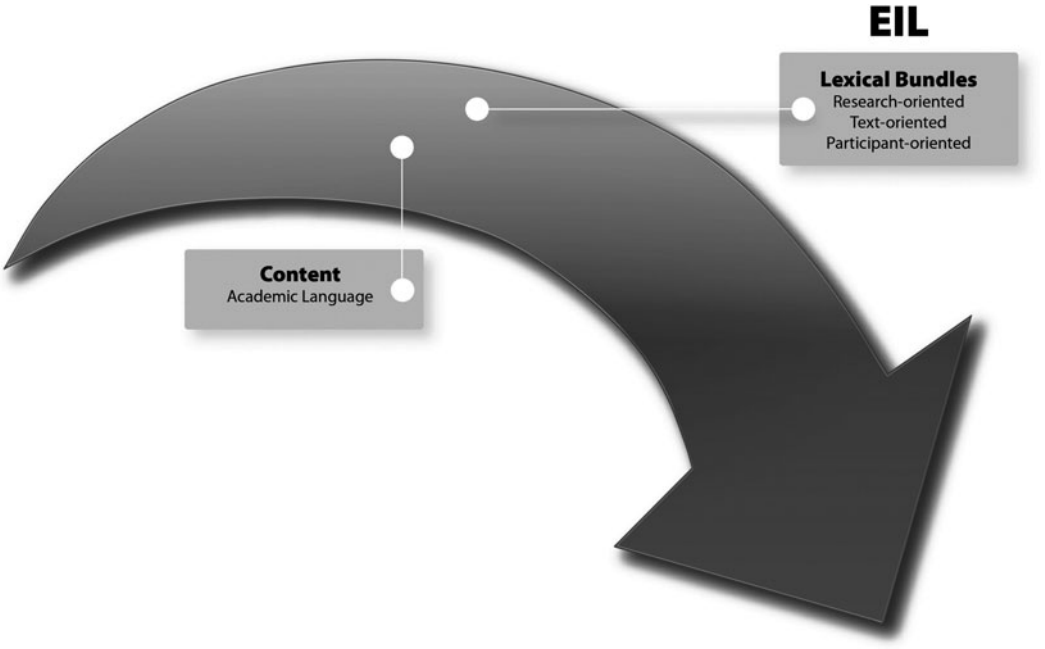


Figure 1. The content-language continuum

https://doi.org/10.1017/S0266078413000151 Published online by Cambridge University Press

experience. In recognizing the very real difficulties of a clear and comprehensive model that will address the complexities outlined above, a shift of focus towards innovation and competitiveness is necessary. Particularly helpful for the task is Ting's Core CLIL Construct (Ting, 2011), which takes a functional view of how the learning process can be enhanced by content-based language instruction: 'CLIL obviously attends to how the learner – not the teacher – is acquiring, using, and mastering the foreign language (FL).' This emphasis on learning rather than teaching is in line with my formulation of a content-language continuum in which discipline-derived language is considered the 'substance' for linguistic interaction and content exchange in academic communities, and not the by-product of language prescription and content formulation.

Since content and language are expected to be dealt with in equal amounts, they should be introduced in two specific moments which I have adapted to my sociocognitive model (see Figure 2): the Language Awareness Phase (LAP) and the Content Awareness Phase (CAP).

These phases introduce a variation of Ting's model of CLIL operand interaction. In this sense, the language awareness phase entails the deployment of specific discourse elements such as register, discourse strategies, connective devices and lexical bundles (Operands 1, 2). The CAP draws on the same discourse elements understood as content-appropriate output (Operand 3). The whole cycle weaves together a coherent interplay of content-language variables resulting in appropriate output (the desired result). However, this model does not seek to describe how and when content-language acquisition takes place from a cognitive-positivist perspective. Instead, it provides insight into the interplay of content and language-driven variables with which learners are expected to interact in the academic field in order to attain literacy.[1] This is important, especially with regard to possible adjustments of the instructional practices applied in contexts where the development of language skills and content instruction is crucial. In other words, the cycle draws on contextually-sensitive content knowledge (teachers' knowledge of the subject matter – language and discipline content) (cf. Shulman, 1987).

LBs' instructional strategy

At the core of the LB-driven instructional practice lies the notion of skill mastery. Thus, successful LB-driven acquisition calls for a multipronged approach that includes direct LB instruction, multiword learning strategies and extensive multi-skills

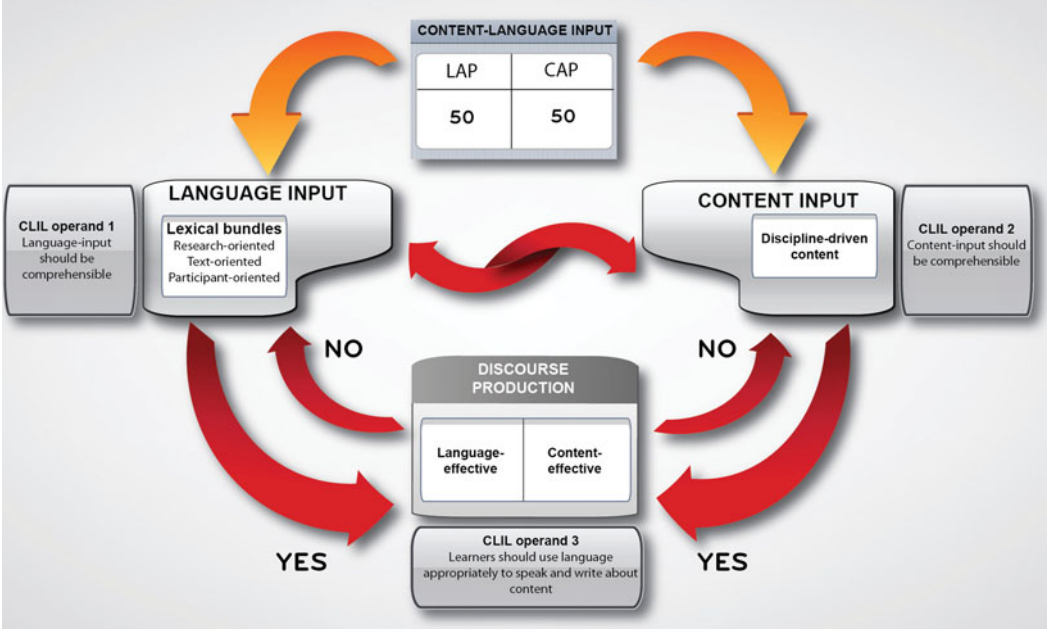


Figure 2. The CLIL cycle

https://doi.org/10.1017/S0266078413000151 Published online by Cambridge University Press

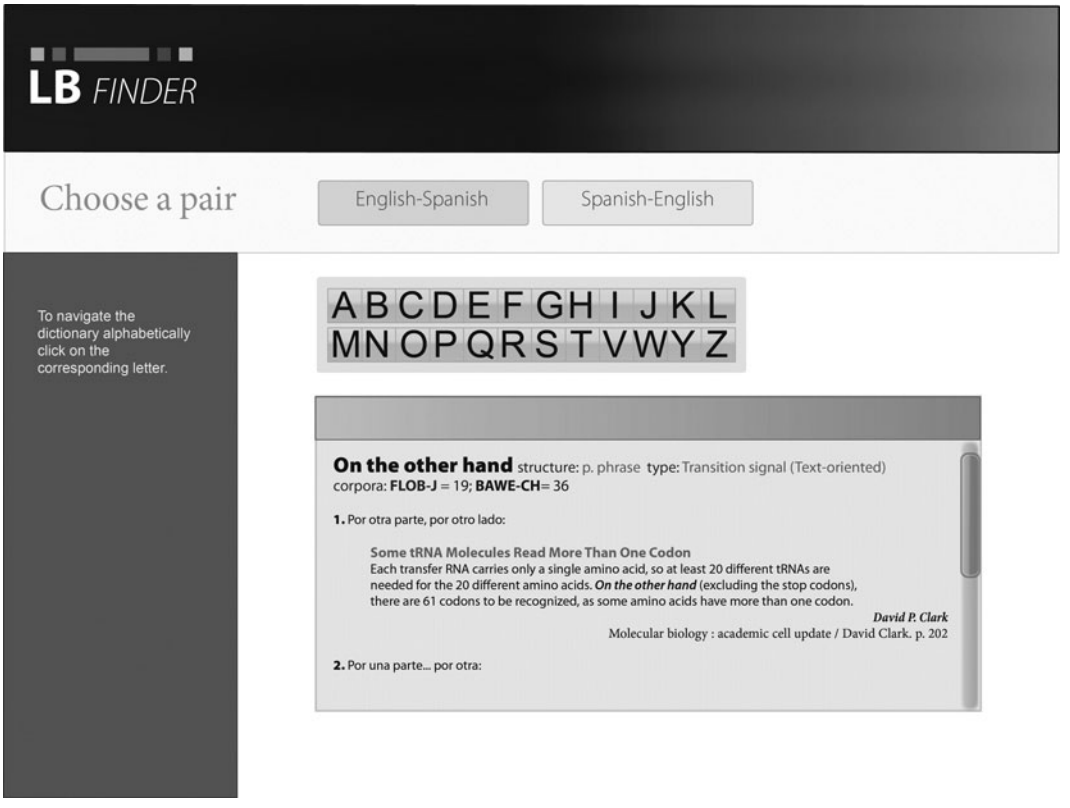


Figure 3. The LB search tool

practice in order to ensure the emergence of a psycholinguistic 'feel of register' conveyed by these chunks, a process which I define as the speakers' ability to effectively engage in register-driven language use. In principle, the LB instructional practice proposed in this article is closer to a socio-cognitive[2] perspective than to a cognitive positivist view. In line with this, I believe that teachers' practice must be adapted depending on the context within which it takes place.

At the most general level, it is reasonable to say that any attempt to define the parameters of any LB-driven instruction within the boundaries of CLIL programme must encompass the following elements:

- In order to change the learners' knowledge and skills, instructional types of courses are recommended (skill-driven, not simply informational).
- A combination of on-line, classroom and real-life learning environments is crucial for the development of both content and language skills.
- This type of instruction should be delivered in both synchronous (application sharing, instant messaging, etc.) and asynchronous (self-paced and collaborative writing tools) manner.
- A combination of rapid and traditional instructional design is recommended.

A LB search tool

In order to contribute to the pedagogical practice of LB instruction, I would like to close my article with a short description of a little tool I have created to help maximize learners' engagement and comprehension of the discursive dimension of LBs in context (written academic texts). (See Figure 3.) For this purpose, I have designed an interactive desktop application drawing on a database of sequences commonly used in disciplines such as biology, engineering and history. These bundles were extracted from the FLOB-J corpus, containing eighty 2,000-word excerpts from published

https://doi.org/10.1017/S0266078413000151 Published online by Cambridge University Press

academic texts, retrieved from journals or book sections, and the BAWE corpus (published in 2008). Both corpora provide a wide range of information concerning the use of LBs in academic writing in disciplines such as life sciences, physical sciences and social sciences. LBs extracted from these corpora are organized in alphabetical order as fully searchable multi-word entries translated into Spanish and with a sample of use in context (a frame). By interacting with the application learners can gain insights into how LBs are used in their respective disciplines while they become aware of the formulaic nature of these sequences. The advantage of the application is that it contains all the required information for the early stages of the instructional process accessible in one place: corpus co-occurrence, pragmatic function plus real-life samples and links to specific online documents. The decision to integrate a Spanish version into the application was a practical one: in order to develop a 'feel of register', and to show students that LBs typical of academic Spanish often do not translate directly into their English equivalents.

Conclusion

In this article, I have tried to show how Ting's Core CLIL Construct, with its emphasis on learner-centredness, combines well with the flow-like character of the CLC to provide insightful theoretical framework for the teaching of LBs in the academic writing classroom. It is proposed that this approach promises to alleviate many of the problems that beset academic English teaching and learning in Colombian universities, and that it provides a more convincing integration of language and content than that provided in traditional content-based programmes. ■

Notes

1 Literacy is defined as 'a broader discourse competence that involves the ability to interpret and critically evaluate a wide variety of written and spoken texts' (Kern, 2000: 2).

2 Regarding this definition Atkinson (2002) points out: 'Neither language acquisition nor language use – nor even cognized linguistic knowledge – can be properly understood without taking into account their fundamental integration into a socially-mediated world.'

References

Atkinson, D. 2002. 'Toward a sociocognitive approach to second language acquisition.' *Modern Language Journal* 86(4), 525–45.

Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. 1999. *Longman Grammar of Spoken and Written English*. Harlow, Essex: Pearson Education.

Conklin, K., & Schmitt, N. 2008. 'Formulaic sequences: are they processed more quickly than nonformulaic language by native and nonnative speakers?' *Applied Linguistics*, 29 (1), 72–89.

Cortes, V. 2004. 'Lexical bundles in published and student disciplinary writing: examples from history and biology.' *English for Specific Purposes* 23, 397–423.

Kern, R. 2000. *Literacy and Language Teaching*. Oxford: Oxford University Press.

Mercer, S. 2012. 'Dispelling the myth of the natural-born linguist.' *ELT Journal* 66(1), 22–9.

—. & Ryan, S. 2010. 'A mindset for EFL: learner beliefs about the role of natural talent.' *ELT Journal* 64(4), 436–44.

Nekrasova, T. M. 2009. 'English L1 and L2 speakers' knowledge of lexical bundles.' *Language Learning* 59(3), 647–86.

Shulman, L. 1987. 'Knowledge and teaching: foundations of the new reform.' *Harvard Educational Review* 57, 1–22.

Ting, Y. L. 2011. 'CLIL. . .not only not immersion but also more than the sum of its parts.' *ELT Journal* 65(3), 314–17.

https://doi.org/10.1017/S0266078413000151 Published online by Cambridge University Press