

A Corpus-based Approach to Content and Language Integrated Learning

Mona Arhire, Mihaela Gheorghe, Doru Talabă

Transilvania University (Romania)

mona.arhire@unitbv.ro, m.gheorghe@unitbv.ro, talaba@unitbv.ro

Abstract

Content and Language Integrated Learning (CLIL) has been gaining increasing ground over the last two decades. In the context of an obvious tendency towards the internationalization of study programmes across Europe, we witness a spread in the implementation of academic study with English as a working language. Therefore, an increased interest in the improvement of content-based language acquisition is justified. In this paper we advocate the usefulness of corpus-based approaches to teaching and learning. Specialized electronic corpora and adjacent tools are deemed to have the potential of offering fast access to reliable texts conceived by native speakers in the respective fields of study. The benefits range between simple searches for specialized terminology to acquiring information in terms of content knowledge, but they are also related to the use and usage of language items. The main advantages of corpora stem from the possibility to obtain contextualized information from authentic and reliable texts, where contextualization offers a structured retrieval of a language item in all the available contexts provided by the corpus.

A package of a multilingual open-ended corpus and electronic tools to enable handy on-line access to the corpus content is believed to be considerably helpful in balancing between content and language acquisition in international higher education. The usefulness of the corpus extends much beyond the area of CLIL so as to ultimately enhance intercultural communication in the priority scientific areas across study programmes in Europe. Corpus-based CLIL is also envisaged to have a formative effect both on non-native English speaking students and trainers. For the former, it stimulates the autonomy of learning outside classroom settings, thereby having a highly motivational effect and fostering cognitive activities. For the latter category, i.e. the educators, reliable and authentic corpora can enhance their linguistic abilities in direct relation with the content knowledge they possess. By means of corpora and electronic tools, translation can also be practiced by non-native lecturers since they resort to it in order to create materials and content lectures in English. The writing of teaching material based on corpora and corpus-based language learning have been suggested before, but we argue that specialized corpora along with some specific tools support CLIL at academic level.

1. Introduction

“Integrating an international dimension in curricula highlights the importance of languages. On the one hand, proficiency in English is *de facto* part of any internationalization strategy for learners, teachers and institutions. [...] On the other hand, multilingualism is a significant European asset: it is highly valued by international students and should be encouraged in teaching and research throughout the higher education curriculum. Additional European language skills widen career prospects...” [9]. One of the steps taken in line with the EU policy above has been the introduction of Content and Language Integrated Learning (CLIL) in many European countries. It is currently undergoing steady development at all the educational levels and increasing interest in the field is displayed by educators and researchers alike. “CLIL is a dual-focused educational approach in which an additional language is used for the learning and teaching of both content and language” [2].

In this paper we address the main purpose of CLIL, namely the internationalization of study. We are attempting at suggesting ways to enhance intercultural communication, multilingualism and employability, envisaging also increased chances for the mobility of teaching staff, students and professionals around Europe. The educational level we focus on is tertiary education where English is the working language, the trainers lecturing in English in addition to their mother tongue. More precisely, we are proposing a methodology meant to optimize the teaching and learning of content knowledge and the way English is handled by non-native speakers of English. The two interrelated facets of CLIL – content and language – raise the requirement of the parties’ proficiency in English, which directly affects the overall effectiveness of the study.

As starting point we have used the assumption that, notwithstanding the fact that both trainers and trainees involved in CLIL are advanced speakers of English, there is room for improvement. The

present approach relies on the belief that such improvement can be obtained by alternating traditional linguistic resources in CLIL-related undertakings by applying corpus-based activities.

2. Preliminaries

One of the considerations triggering our approach relates to findings deriving from studies which conveyed evidence of the fact that lecturers involved in CLIL tend to use traditional linguistic resources to prepare their lectures. To refer but to a study undertaken in Spain, out of the alternatives given in a questionnaire to lecturers involved in Spanish CLIL (dictionaries, glossaries, machine translation and web pages), dictionaries resulted to have the upper hand, while web pages are second in range [5]. The study also revealed that all the trainers subject to the investigation constantly resort to translation from Spanish into English when preparing their lectures and in-class presentations. Given that the validity of these findings is limited to the Spanish environment, for the time being, we can only hypothesize that they can be extended to most of the European tertiary education systems in which CLIL is gaining ground. As can be noticed, the use of corpora is never appealed to in this study, although corpora exhibit clear advantages as compared to dictionaries.

Nevertheless, corpora started to be integrated in CLIL environments. In Italy, for instance, corpus-driven data and corpus-tailored activities were successfully explored in terms of both content and language acquisition [1].

Considering the methodological wealth that can be derived from corpus linguistics, our belief is that electronic corpora can substantially aid the teaching and learning process in CLIL environments. Specialized text corpora can be successfully adopted to enhance the trainers dealing with the lectures they deliver and the students' comprehension thereof. While hypothesizing that the European countries face similar language-related conditions in CLIL, a multinational endeavor could be suitable to test this hypothesis and ultimately to head for a common and more unitary approach to optimizing CLIL by incorporating corpus methodology in the curricula at European level.

3. Advantages of corpora in CLIL

"A corpus is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research" [6]. The corpus is the study object of Corpus Linguistics, a study area that has exceeded its limits so as to intertwine with branches within linguistics and beyond it. During the last decades, its interdisciplinary dimension has extended to generate results deriving from large-scale research. Especially since the beginning of the cooperation with Information Technology (IT) and the emergence of electronic corpora, corpus linguistics has established partnerships with sundry theoretical and applied study areas. Thereby a die was cast in various fields for which corpus methodology is nowadays indispensable or highly valuable.

Corpora have been used for years now for educational purposes in second and foreign language acquisition, but, to the best of our knowledge, less so in CLIL. Here, the primary aim is not the acquisition of a language as such, but the learning and teaching process is targeted at students' gaining content knowledge in a foreign language which they possess at advanced level. Therefore, corpora will operate differently in the specific CLIL environment.

The main benefit that CLIL can derive from corpora stems from the possibility to view language in larger stretches and thus retrieve contextualized information of various kinds. This can be partly achieved with the help of electronic tools, named concordancers. The reason why corpora are preferred to dictionaries for terminology clarification is that corpora exhibit the search term in as many contexts as are available in the corpus. But corpora seem to be even more valuable for the retrieval of information about language use and usage than about specialized terminology. The authentic use of words or lexical clusters is illustrated at its best when enlisted in several concordance lines. For example, in the case of the sentence extract: *...their efforts to prevent such incidents to happen again*, the key words *prevent*, *incidents* and *happen* have been searched for together to check their collocational validity. The corpus revealed the versions *prevent an incident happening* and *prevent an incident from happening*, but not *prevent an incident to happen* [3].

Corpora can also clarify ambiguity and indicate semantic differences when different prepositions join the same verb or noun and can reveal grammatical information relative, for instance, to the countability of nouns, agreement, contrastive structures, etc. The possibility of clarifying between false friends has also been demonstrated [4]. Synonymy and polysemy actually entail a choice of the right lexical items, whose use is best illustrated in their co-text. These, as well as many other language data, would be otherwise obtained from several various dictionaries or reference books with less effectiveness and more time consumption.

Automatic frequency lists can also be helpful. For instance, the difference between *There are not any ...* and *There are no ...* has been verified in a corpus by means of a frequency list. The results showed that the latter version is much more frequent (therefore preferred) than the former, entailing also a syntactic and cohesive difference [3].

What is more, the compilation of a corpus, hard as it might be, is worthwhile since it can constitute the root for a plurality of investigation purposes beyond the initial ones.

4. The corpus type for CLIL in specialized academic study

The corpus that we envisage for CLIL at BA or MA level is first of all a multilingual one, covering progressively as many European languages as possible. The English sub-corpus is the core component since English is the working language. The other sub-corpora will first and foremost address the needs of individual cultures corresponding to the respective mother languages. Then, access to the sub-corpora in other languages will enable the acquisition of additional European languages once mobility or employability offers become options within the European community. Therefore, the corpus is a multilingual comparable one, comprising specialized texts with similar features in similar domains. One of the most important general characteristic of corpora is their reliability. A corpus derives its reliability from several aspects, some of which will be further touched upon with direct reference to the corpus under discussion. A chief condition for corpus reliability is granted by the authenticity of the texts it includes, as well as by the careful choice of authorship. In this respect, the corpus under discussion will be made up only of authentic scholarly work, manuals, coursebooks, *etc.* written by native speakers of the language they represent. Thereby, both the quality of the content and that of the languages involved are safely assured. Additionally, the size of the corpus is of utmost importance. The more sizable a corpus is, the more relevant data it will generate. Isolated search results might not be relevant and can be contradicted if a more sizable corpus is used for data retrieval. Last, but not least, the corpus is machine held. The electronic format is just an implicit condition nowadays since it is the automatic data processing that enables users to obtain substantial information about search items. No manual analysis could ever compensate for the advantages that electronic corpora can provide when accessed by means of specific electronic tools. As for accessibility, such a corpus should be freely on-line accessible to the entire European community and it should also be an open, monitor corpus to be constantly added and updated. Other features of the corpus, such as user-friendliness, timeliness and ethical issues are to be taken into account.

5. Conclusions

The specific benefits of corpora for trainers lie in the possibility to acquire translation skills and to prepare their lectures and in-class presentations, to access documentation material for lecture delivery and research, as well as to improve their English. Trainees are themselves advantaged by higher quality lectures, but, more importantly, the corpus lends itself not only to in-class applications, but also to individual ones, thus fostering autonomous learning outside the classroom setting and beyond the time-limited institutional instruction schedule. Learner autonomy is an important part of the CLIL approach: "CLIL as a fused subject provides a learning environment which promotes the capacity for self-organization" [7]. The corpus-based approach enables students to mine language descriptions in a self-directed way [8] and to develop their reading and writing skills, while understanding how languages – either individually or contrastively – are used in the particular register they deal in. The acquisition of other European languages represents an additional kind of openness for them and heightens their possibility to move freely and effectively in the European space, increasing their chances for employability, exchange of experience, multilingualism and multiculturalism, and helps them to become responsible global citizens.

Despite our assumption that corpus-based activities can considerably enhance CLIL, we do acknowledge that specific pre-tests in CLIL settings in several countries are necessary before concluding on the specific corpus-based methodology to be developed and on the refinement of the electronic tools. Therefore, a multinational cooperation is hereby called for to include countries where CLIL is already operational with a view to stimulating other countries to embrace the methodology.

References

- [1] Carloni, G. (2013). Integrating English corpus linguistics and online learning environments at university level. <http://www.ijitcd.com>.
- Coyle, D., Hood, P., Marsh, D. (2010). *CLIL Content and Language Integrated Learning*, Cambridge, Cambridge University Press.
- [3] Hunston, S. (2014). *Corpora in Applied Linguistics*, Cambridge: CUP.

- [4] Imre, A. (2013). Traps of Translation. A Practical Guide for Translators. Brasov: Transilvania University Press.
- [5] Riera, J.B. and Arevalo, C. M. (2013). CLIL and translation in tertiary education in *Revista de Lenguas para Fines Especificacion*, 19, Spain: Servicio de publicaciones de la ULPGC, pp. 29-59.
- [6] Sinclair, J. (2005). Corpus and Text - Basic Principles. In M. Wynne (ed.) *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books, pp. 1-16.
- [7] Wolff, D. (2011). CLIL and Learner Autonomy: relating two educational concepts, *Education et Sociétés Plurilingues*, 30 June 2011, <http://www.cebip.com/datapage.asp?l=1&id=>.
- [8] Woolard, G. (2000). Collocation-encouraging learner independence. In M. Lewis (ed.). *Teaching collocation: Further development in the lexical approach*. Hove, England: Language Teaching Publications, pp. 28-46.
- [9] <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52013DC0499#text>