



## Backbone

**Michel Meuret**

E-mail: [mmeuret@jura.cci.fr](mailto:mmeuret@jura.cci.fr)

Chambre de Commerce et d'Industrie du Jura (France)

### Abstract

*This project addresses language teachers and teacher educators in secondary, higher and vocational education with a focus on content and language integrated learning (CLIL) and e-learning.*

*The overall project objective is to offer language teachers and teacher educators in CLIL settings pedagogic corpus and e-learning solutions involving authentic, subject-specific spoken interviews. With its "do-it-yourself" orientation, BACKBONE seeks to provide for lesser taught languages as well as regional, socio-cultural and subject-related varieties.*

*More specific project objectives include*

- *creation of pedagogically annotated and enriched web-based corpora of video-recorded interviews in English, French, German, Polish, Spanish, Turkish and English as a Lingua Franca;*
- *development of pedagogically designed tools for corpus transcription, annotation and search;*
- *pedagogic evaluation of the BACKBONE approach, corpora and tools in Moodle-enhanced CLIL courses in secondary, higher and vocational education;*
- *dissemination of the BACKBONE results and outcomes with a special focus on online dissemination support and a European BACKBONE conference;*
- *exploitation of the BACKBONE results and outcomes through teacher training workshops and an online exploitation service.*

*The project consortium consists of 8 partners from 7 European countries: France, Germany, Ireland, Poland, Spain, Turkey and UK: 6 HE institutions, 1 VE institution and 1 consultancy company. Together they represent CLIL-specific language teaching needs and pedagogical experience, linguistic and computational know-how, pedagogic evaluation competence, as well as dissemination and marketing expertise.*

*The methodological approach combines desk studies and empirical research, software and website development, e-learning design and blended language learning (BLL) implementation, corpus compilation, evaluation studies, workshop and conference organisation.*

*BACKBONE : EACEA Key Activities 2 programme (languages) – 2009-2010*

143502-2008-LLP-DE-KA2-KA2MP <http://www.uni-tuebingen.de/backbone>

### Introduction

Pedagogically neglected languages in Europe include lesser taught languages, regional & socio-cultural varieties of more frequently taught languages, and non-native speaker varieties of lingua franca languages. A dramatic lack of pedagogical inclusion with regard to these varieties can be witnessed across all educational sectors; this is even reinforced by a lack of suitable published learning materials.



To meet these challenges, teachers are required to be more autonomous, in particular with regard to the availability of thematic contents relevant for language learning. In this respect, they find themselves in the same boat with teachers in bilingual subject classes or vocational courses. The methodological challenges are the same: how is it possible to support content and language integrated learning (CLIL) in its various manifestations and thus to make use of the motivating potential of language learning in authentic contexts? Important synergy effects are created when CLIL is combined with e-learning. Considering the power of the internet for providing rich content resources as well as supporting communication and collaboration, e-learning seems to be a natural environment for CLIL-based activities.

The European Lifelong Learning project BACKBONE (2009-2010) adopts this perspective and proposes a pedagogically motivated corpus approach that enables teachers to collect and exploit spoken discourse resources for teaching pedagogically neglected languages and varieties. The BACKBONE pedagogic corpus approach is based on the constructivist principles of learner and teacher autonomy, authenticated learning, and collaboration; it combines spoken and video-recorded corpus resources geared to linguistically and thematically specialised CLIL purposes with Moodle-based language practice and communicative interaction in blended learning courses (cf. Kohn 2009).

Forerunners of the BACKBONE pedagogic corpus approach are the ELISA corpus (cf. Braun 2005, 2006, 2007 - see <http://www.uni-tuebingen.de/elisa>) and the SACODEYL corpora (cf. Hoffstaedter & Kohn 2009; Pérez-Paredes & Alcaraz-Calero 2009; Widmann, Kohn & Ziai 2008). See: <http://sacodeyl.inf.um.es/sacodeyl-search2>

## Objectives and approach

BACKBONE develops, evaluates, disseminates and exploits a "do it yourself" pedagogic corpus approach that empowers teachers to collaborate in the creation and pedagogical deployment of authentic content for web-based language learning and teaching in specialised CLIL contexts. The "engine" of this approach is an ensemble of corpus and e-learning tools that are developed and integrated to support pedagogic corpus annotation and enrichment, online pedagogic corpus search, and embedding of corpus-based learning materials and activities in Moodle courses. BACKBONE tool development builds on products available from the European Minerva project SACODEYL (2005-08) - see <http://www.um.es/sacodeyl>

Content-wise, BACKBONE addresses the constraints and needs of pedagogically neglected languages & varieties. It is applied to 7 languages representing 3 areas of neglect: "lesser taught" (Polish, Turkish), "ignored non-standard" (regional and socio-cultural varieties of English, French, German, and Spanish), and "non-native speaker lingua franca" (ELF: English as a lingua franca). For each of these languages and varieties, a corpus of video-recorded spoken interviews with speakers from different walks of life (e.g. occupation, social class, region, dialect) is compiled. (50 interviews of 10 minutes each for English and ELF, 25 interviews of 10 minutes each for French, German, Polish, Spanish and Turkish). Backbone involves 8 partners from 7 European countries: France, Germany, Ireland, Poland, Spain, Turkey and UK. It is coordinated by Tübingen University, Applied English Linguistics.

To explore the pedagogical potential of the BACKBONE corpora, piloting courses are implemented in 5 language learning settings: regular foreign language classes in secondary and/or higher education integrating culture and language learning; bilingual subject classes (e.g. biology, geography, history) in secondary education with incidental foreign language learning; LSP courses in higher education integrating foreign language and special subjects; community interpreter courses in higher education; and vocational foreign language classes in further education.

## **Pedagogic corpus design, creation, and search**

In spite of the widespread agreement in the research and teaching communities that corpora are in principle a precious resource for language learning and that they should be made accessible to teachers and learners alike, there are still obstacles that prevent the effective pedagogical use of corpora, especially their direct use by learners. One of these is perhaps temporary and has to do with teacher training.

Recent surveys suggest that many teachers simply do not know (enough) about corpora to use them appropriately and confidently with their learners.

The lack of knowledge about corpus-based learning among teachers is compounded by the general and omnipresent problems with the use of technology 'in the classroom'.

Our decision for pedagogic mediation has far-reaching implications throughout all levels of corpus design and creation. Pedagogic mediation begins with the actual interview recordings, in particular the selection of speakers, topics, and interview set-up settings. For the BACKBONE interviews, we analysed language course programmes and materials in secondary, higher and vocational education to identify a range of relevant, CLIL-specific themes. We also talked to teachers who would later be interested in taking part in the pedagogic evaluation.

The interview approach is not a sophisticated one. The main purpose is to get the interviewee to relax and talk; conversational interaction is not in the foreground. Typical topic areas include culture, economy, urban and rural life, social issues, health and social security, education, environment, government and politics.

Since the interviews are intended for learning contexts, transcription uses an orthographical notation including punctuation, albeit slightly adapted to rendering spoken discourse. A few pre-defined mark-up codes are used to specify, e.g., breaks, truncations, alternatives, or comments. Fillers, repetitions, and hesitation phenomena are only accounted for if considered to be meaningful. These transcription conventions are clearly influenced by the pedagogic purpose of the entire corpus approach.

Annotation applies to short transcript sections, which are annotated in a drag & drop fashion with regard to properties deemed relevant for language learning purposes by the annotator-teacher. The annotation categories chosen refer to thematic, grammatical, lexical, and textual characteristics, as well as CEF level specifications. The aim of our annotation is not a classificatory one. The categories are rather meant to support pedagogic search; they can be freely defined and thus tailored to capture an individual corpus's pedagogic potential.

In addition, enrichment resources can be managed in a Virtual Resource Pool (VRP) and linked to transcript sections. In the case of the BACKBONE corpora, this includes small corpus-based learning modules created in Telos Language Partner (cf. Kohn 2008) as well as instructions for exploratory, communicative and collaborative learning activities.

The BACKBONE interviews are available in an online search interface offering 4 pedagogically motivated search modes: In 'Browse', entire interviews can be viewed and listened to, which facilitates contextualisation and discourse authentication. 'Section search' presents the corpus's annotation category tree and is used to search for individual interview sections that comply with a specified combination of thematic and linguistic annotation categories; words and phrases that satisfy one of these categories can be highlighted. 'Word search' produces concordances; and 'Co-occurrence' lists sections that contain a number of specified words in free distribution. These two lexical search modes can be combined with selecting annotation categories, thereby limiting the search scope to sections that fall, e.g., in a certain topic area, or belong to a preferred CEF level.

## **The backbone corpora**

### **Topic Oriented Learning Activities**

Topic oriented learning activities include activities focussing on listening/reading comprehension supported by multimedia learning modules, topic driven explorations of the of the corpora, and topic and task driven exploratory internet research activities (WebQuests) using the corpus materials as an opener or starting point.

### **Comprehension**

Multimedia learning modules can help to enhance a section of the video interview with comprehension support. The learning modules should offer pedagogically structured learning activities that help students to work with the interviews in particular with regard to the challenge of understanding authentic spoken discourse. This is especially helpful for learners at lower proficiency levels, but it may also encourage advanced students to listen to an interview section more carefully.

### **Topic Exploration**

Students can use the corpus to explore what different speakers say about a specific topic. They can search for interview sections on specific topics by using the available topic categories. The search can be narrowed down by using the full-text search option of the 'Co-Occurrence Search'.

Explorative activity should be task driven and can be combined with communicative or cooperative activities where students exchange ideas and results of their corpus search with other students, or where they present their results in class or in a Wiki on an online platform.

### **Topic Related Webquests**

Interview sections can also serve as an opener for a guided internet research (WebQuest) on a topic addressed in an interview. It is important to think of authentic and manageable tasks, to provide guiding questions and links to suitable websites. It is very important to make sure that students involved in a WebQuest do not use the internet for just copying and pasting information but that they really use the provided information to get a deeper insight into a specific matter.

### **Vocabulary Development**

Activities in certain content areas or topics can also be combined with lexical activities. The BACKBONE search tool offers different search options that can be used for activities aiming at vocabulary development, concordance-search and co-occurrence search.

The search tool offers a powerful concordancer that is able to left-sort or right-sort the results. Users can enforce case-sensitivity and they can define the co-text length from 8 to 15 words. Moreover, they can limit the search to specific sections by using the topics and the other annotation categories as further filters.

## **Grammar**



The Backbone search tools offers the option to combine different search categories. You can, for instance, search for a certain topic (e.g. “presentation”) and grammatical features (e.g. Prepositions). The search result can then, for instance, be used to create a learning activity combining a listening with a grammar exercise

## **Communicative Functions**

If a corpus is annotated with regard to communicative functions, teachers and students can easily see search for interview sections representing certain functions like “giving opinions”. Students can, for instance, in a first step use the corpus search to explore how speakers express their opinion on a certain topic. This is an exploratory task that should be followed by a communicative activity where students then have to express their own options discussion a certain topic in a forum. If the communicative functions are not annotated in a corpus, students can also use the co-occurrence search and enter function specific word, e.g. “think” to search for corresponding interview sections.

## **Spoken Discourse And Regional Or Social Varieties**

Spoken language and regional or social varieties pose a particular challenge to language learners.

Learning activities should, therefore, offer comprehension support and make students aware of lexical or grammatical characteristics of spoken language and regional or social varieties.

Beyond design, creation and search, the pedagogic orientation of our BACKBONE corpora also extends to their deployment in learning tasks and activities. Guiding principle is the pedagogic integration of focus on form with focus on collaborative communication. BACKBONE combines Telos Language Partner and Moodle to support these two complementary task and activity types.

It is quite evident that corpus data have a natural affinity to focus on form. A decisive advantage of the BACKBONE approach, however, is that it enables looking at grammatical or lexical form within a specified thematic area, e.g. tenses or collocations in the area of fishing. Note that in our context, the purpose of this content/language combination is not a linguistic-descriptive but a pedagogic one, i.e. the CLIL-specific integration of content learning with language learning. Suitable learning tasks are made available through Telos modules, which are integrated and made searchable as enrichment resources.

Embedding Telos learning modules in Moodle courses opens up flexible possibilities for pedagogically combining corpus-based focus-on-form activities with online collaboration and communication. The forum, chat, and wiki activities provided by Moodle can be easily enhanced through linking to externally available web 2 applications – from e.g. WikiSpaces to YouTube to Skype. Starting from corpus data, it thus becomes possible for learners to continue to engage in authentic online discourse events and learning experiences.

## **Conclusion**

We are convinced that the BACKBONE approach opens up an innovative potential for authenticated language learning. It must not be overlooked, however, that teachers play an important part in bringing this potential to life. In this connection, a preliminary observation from previous pedagogic evaluation studies with ELISA and SACODEYL corpora in German secondary schools is of interest. It concerns a sometimes sharp discrepancy between the initial enthusiasm with which teachers welcome the new pedagogic opportunities and the often rather poor uptake of these opportunities in their daily school practice.

The occasional e-teaching workshop is not enough to close this gap. What is needed are in-depth pedagogic implementation studies that combine and integrate two key measures: first, long-term evaluation of the pedagogic corpus approach in real courses that are part of the regular school programme; second, continuous teacher training and support in direct connection with these courses.

Another condition for teacher autonomy in e-learning contexts is worth mentioning: successful e-learning requires e-teaching outside and beyond the classroom – and it requires work regulations that make this possible.

**The partners gratefully acknowledge the funding provided by the European Commission. This publication reflects the views only of the authors, and the Commission cannot be held responsible for any use which may be made of the information contained therein.**

## References

- Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17/1, 47-64.
- Braun, S. (2006). ELISA – a pedagogically enriched corpus for language learning purposes. In S. Braun, K. Kohn & J. Mukherjee (Eds.), *Corpus Technology and Language Pedagogy. New Resources, New Tools, New Methods* (pp. 25-47). Frankfurt/M.: Peter Lang.
- Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL*, 19, 307-328..
- Hoffstaedter, P. & K. Kohn (2009). Real language and relevant language learning activities: insights from the SACODEYL project.
- Kohn, K. (2008). *Telos Language Partner: DIY authoring for content-based language learning.*
- Pérez-Paredes, P. & J.M. Alcaraz-Calero (2009). Developing annotation solutions for online data driven learning. *ReCALL* 21/1, 55-75.
- Widmann, J., Kohn, K. & R. Ziai (2008). The SACODEYL search tool: exploiting corpora for language learning purposes. In Frankenburg-Garcia, A. (Ed.), *Proceedings of the 8th Teaching and Language Corpora Conference, 3-6 July 2008, Lisbon* (pp. 321-327). Lisbon: Associação de Estudos e de Investigação Científica do ISLA-Lisboa.
- Kurt Kohn, Petra Hoffstaedter, Johannes Widmann. *BACKBONE – Pedagogic Corpora for Content & Language Integrated Learning – Eurocall conference 2009*