

Project

Pedagogic Corpora for Content and Language Integrated Learning. Insights from the BACKBONE Project

Kurt Kohn
Applied English Linguistics, University of Tübingen (Germany)

Kurt.kohn@uni-tuebingen.de

Abstract

BACKBONE is a European LLP/Languages project (1) (Jan 2009 - Feb 2011), whose overall objective is to provide foreign language teachers in CLIL settings with innovative language learning solutions. To achieve this goal, pedagogic corpora of spoken interviews are combined with corpus-related e-learning activities in blended learning scenarios. The seven BACKBONE corpora contain video interviews in English, German, French, Polish, Spanish and Turkish as well as in European manifestations of English as a Lingua Franca (ELF). The interviews have been transcribed and pedagogically annotated with regard to thematic and linguistic features; additional enrichment resources include ready-made language learning modules as well as suggestions and instructions for exploratory and communicative learning activities. The BACKBONE search interface provides free online access to the interviews and enrichment resources. It supports pedagogically motivated searches using thematic and linguistic categories as well as lexical searches with words and phrases. 24 CLIL-related pilot courses have been implemented and evaluated in secondary, higher and vocational education; they demonstrate how BACKBONE search results can be used to facilitate individual and collaborative learning in Moodle-based blended learning activities.

A suite of pedagogic corpus tools covering transcription, annotation, management of enrichment resources and corpus search is available under a GNU General Public License. The customization and flexibility these tools offer enables teachers to cater to diverse language learning and teaching needs in CLIL contexts or in connection with lesser taught languages and varieties. To facilitate exploitation, the BACKBONE website serves as a 'one-stop-shop' for an ensemble of teacher support facilities including web support for the development and hosting of "guest" corpora and courses.

Key words: Pedagogic corpora, corpus-based language learning, CLIL, Blended Language Learning, e-learning, Moodle.

1. Rationale

Across Europe, school subjects like history and biology are increasingly taught not in the country's native language, but in a second language, frequently English. This approach is known as bilingual education and forms part of the pedagogic concept of Content and Language Integrated Learning (CLIL). In this form of teaching, the main focus is on the subject at hand; the second language is merely used as the medium of instruction. As a result, students practise and learn the second language in a relevant thematic context and through 'real' communication. CLIL interactions in school, however, are not limited to bilingual subject teaching; they can also be found in regular foreign language teaching contexts when the focus of learning is on, e.g., culture, geography or

literature. The CLIL method holds huge potential for language learning – both in its strong version in bilingual subject classes and in its weak version in subject-focused units in regular foreign language teaching: CLIL can help students improve their general language proficiency, become more confident in using the foreign language for communication purposes, and boost their communication skills in relation to specific subject areas.

Success in CLIL is based on two main factors: rich opportunities for communicative interaction and availability of relevant content with which to work. Both factors are closely related to the communicative-constructivist requirements of authenticity, autonomy and collaboration; and they can be significantly enhanced by a blended learning setting combining web-based corpus resources with tools for computer-mediated communication (CMC).

Written and spoken text corpora were first developed for purposes of linguistic description, e.g. the British National Corpus (BNC) or the International Corpus of English (ICE), and to support the production of dictionaries (e.g. Sinclair 1987) and grammars (e.g. Sinclair 1990; Biber et al. 1999). Extensive deployment of corpora and corpus techniques (e.g. concordances, word lists frequency counts) for language learning and teaching purposes, in particular data-driven learning, became possible with advancements in PC and internet/web technology (Aston, Bernadini & Stewart 2005; Boulton 2011; Braun, Kohn & Mukherjee 2006; Flowerdew 2009; Johns & King 1991; Sinclair 2004; Tribble & Jones 1990; Wichman, Fligelstone, McEnery & Knowles 1997). In the wake of this pedagogic turn, the range of corpus types has been extended to include e.g. non-native speaker corpora of written and spoken learner English (ICLE and LINDSEI; see Granger, Dagneaux and Meunier 2002) or small, do-it-yourself corpora focusing on genres and topics of immediate relevance to a specific group of learners (Tribble 1997; Aston 2002). A more recent development concerns the change from a primarily descriptive corpus concept to a pedagogical one. An early prototype is *ELISA*, a small interview corpus of everyday spoken English developed by Sabine Braun (2005, 2007). Incorporating Widdowson's (1991, 2003: 102ff.) principle of pedagogical mediation, *ELISA* adopts a consistent pedagogical conceptualization of the entire corpus process from compilation to annotation, enrichment and search. The overall objective is to help teachers and learners proceed from decontextualized textual data to context-embedded discourse interaction and thus to facilitate and promote learner authentication (Widdowson 1998, 2003: chap.8). *ELISA*'s pedagogical approach was adopted and further developed, both in terms of design and corpus tools, by the European Minerva project *SACODEYL* (2005-08) (Braun 2010; Hoffstaedter & Kohn 2009; Pérez-Paredes & Alcaraz-Calero 2009; Pérez-Paredes 2010; Widmann, Kohn & Ziai 2010).

Corpus resources and tools can be adapted to function as an important content-related e-learning pillar of CLIL; the second pillar, communication and collaboration, is provided by internet and web2 technologies: forum, chat, skype, wiki or blog are some of the key tools that help facilitate communicative and social contact as well as collaborative production (Guth & Helm 2010; Kohn & Warth 2011). The pedagogical integration of these two dimensions of e-learning, access to online resources and communicative interaction, creates a huge potential for authenticated foreign language learning through CLIL collaboration. Relevant language learning activities include, for example, explorations of linguistic means of expression in a corpus or in Google, forum discussions of a subject or language-related topic, collaborative creation of multimedia wiki documents, conversations in Skype, or blog entries reflecting on learning challenges and strategies. This e-learning extension of CLIL offers flexible opportunities for authenticated written and spoken production. It is thus particularly suitable for ensuring a more balanced distribution of reception and production activities than is usually possible in a face-to-face classroom setting (2).

E-learning integration of corpus resources with the communication and collaboration facilities of the internet and web2 can be exploited in different learning contexts. The range includes the ordinary foreign language classroom, bilingual subject teaching in its

various manifestations, or incidental language acquisition through communicative contact in the social web.

In the following chapters, I will describe the pedagogical corpus approach developed in the European LLP/Languages project "BACKBONE – Corpora for Content and Language Integrated Learning" (Jan 2009 – Feb 2011). A strong emphasis is on pedagogical corpus principles, tools and contents. In addition, Moodle-based pilot courses in secondary, tertiary and vocational education demonstrate how corpus resources can be combined with online communication and collaboration in blended learning settings to enable learners and teachers to engage in rich language and content integrating learning experiences (Kohn, Hoffstaedter & Widmann 2010).

2. Objectives and methodological approach

The BACKBONE project addresses the pedagogic content needs and challenges of language teachers in secondary, higher and vocational education with regard to a pedagogic integration of CLIL and e-learning. The overall project objective is to offer do-it-yourself e-learning solutions based on a pedagogic corpus approach involving spoken interviews on a wide range of topics. Special attention is given to lesser taught languages, to regional, socio-cultural and subject-related varieties of more frequently taught languages, as well as to English as a Lingua Franca.

More specific project objectives concern pedagogic research, pedagogic corpus tool development, e-learning implementation, pedagogic corpus creation, pedagogic piloting and evaluation, dissemination and exploitation. This involves in particular:

1. carrying out an empirical "fields & needs" analysis of pedagogic scenarios, learning objectives and CLIL topics, a research study on corpus-enhanced language learning & teaching, as well as pedagogic evaluation studies of the BACKBONE corpus tools, corpora and courses [→ pedagogic research];
2. consolidating and extending existing open source corpus tools for interview transcription, collaborative corpus annotation and management, management of enrichment resources, and online search [→ pedagogic corpus tool development];
3. designing and implementing a Moodle platform combining a course area with an area for continuous teacher support [→ e-learning implementation];
4. compiling, annotating and enriching web-based corpora of video-recorded interviews in English, French, German, Polish, Spanish, Turkish, and English as a Lingua Franca [→ pedagogic corpus creation];
5. designing Moodle-based pilot courses for evaluating and exploring the pedagogic potential of the BACKBONE approach in various manifestations of CLIL in secondary, higher and vocational education [→ pedagogic piloting & evaluation];
6. promoting the BACKBONE approach and outcomes with an emphasis on awareness raising and distribution of the BACKBONE corpora in secondary, higher and vocational education [→ dissemination];
7. ensuring long-term impact and sustainability in the educational "market" through teacher training workshops in European countries and implementation of an online BACKBONE service [→ exploitation].

Primary users of the BACKBONE approach are learners and teachers in CLIL settings in secondary, higher and vocational education; secondary users are teacher educators in these areas. Both groups of users have been directly involved in the project through the

pilot courses, which are monitored by project partners specializing in CLIL-related language learning and teaching as well as in language teacher education.

The targeted impact and benefit is twofold: Language learners experience the motivating potential of corpus-based content and language integrated learning (CLIL) – they explore the possibilities of e-learning (in blended learning contexts) with regard to learner autonomy, authenticity and collaboration; and they are enabled to further develop their ICT/media competences. Language teachers are given the opportunity to extend their pedagogical and technological competences. Continuous teacher support is implemented to ensure seamless integration of BACKBONE activities with the individual teacher's established teaching approach. Additionally, teachers and educational institutions are invited to deploy the BACKBONE tools in a do-it-yourself fashion to create their own customized BACKBONE corpora and courses.

In BACKBONE, all decisions regarding the methodological approach are consistently influenced by research-based pedagogical considerations. Because of this orientation, input from pedagogic "fields & needs" analyses and background research reports about the theoretical-methodological foundations of corpus-based language learning and teaching plays a key role in the overall BACKBONE approach (also see Braun 2005, 2007, 2010; Hoffstaedter & Kohn 2009; Kohn 2009). Informed by these studies, BACKBONE proposes a do-it-yourself corpus approach that empowers teachers to collaborate in the creation and pedagogical deployment of spoken interviews for web-based language learning and teaching in specialised CLIL contexts. Content-wise, BACKBONE addresses the constraints and needs of pedagogically disadvantaged languages and varieties in CLIL settings. The approach is applied to seven languages representing three areas of disadvantage: lesser taught languages (Polish, Turkish), regional and socio-cultural manifestations of more frequently taught languages (English, French, German, and Spanish), and English as a Lingua Franca (ELF). For each of these areas, a corpus of video-recorded, spoken interviews with speakers from different walks of life (e.g. occupation, social class, region, dialect) is compiled.

The pedagogical orientation has far-reaching implications throughout all levels of corpus design and creation. This begins with the actual corpus recordings, in particular the specification of the communication genre, the choice of topics and the selection of speakers. In terms of communication genre, BACKBONE focuses on interviews; topics were chosen in accordance with the thematic orientation of language course programmes and materials in secondary, higher and vocational education. To increase BACKBONE's application potential, we also talked to teachers who would later be interested in taking part in the pedagogic evaluation.

In keeping with the do-it-yourself approach, the preferred interview recording procedure is not a sophisticated one. The main purpose is to get the interviewee to relax and talk; conversational interaction is not primary. As a result of this, the interviews have a distinct monological, "native narrative" character. The questions are rather short, with the main purpose of encouraging longer descriptions, explanations and opinions; dialogical interaction is thus less frequent. This clearly results in a certain restriction of the range of communicative activities and functions; at the same time, however, this interview approach helps to create fairly natural conditions for the recording of spoken language. In addition, monological interview utterances are communicatively far more interesting than it might appear at first sight. The ability to describe, explain and evaluate things in an interview is also of crucial importance for dialogical communication. Speakers who are not able to express themselves in a monological interview will undoubtedly have serious problems in conversation. Learning with interview-based material should thus be given a key role in preparatory communication tasks. Nevertheless, it must be conceded that conversation-specific interaction and related means of expression are not covered in the BACKBONE narratives. With an adapted elicitation procedure, however, the BACKBONE approach could be easily extended to capturing conversation-style interactions as well.

The pedagogical orientation also thoroughly influences corpus tool development. The “engine” of the BACKBONE corpus approach is an ensemble of corpus tools created and adapted to support pedagogic corpus annotation and enrichment, as well as online corpus search. Key features are designed with a pedagogic purpose in mind and thus differ from descriptive corpus tools in important respects. BACKBONE tool development builds on open source products available from SACODEYL. Development in BACKBONE focuses on further consolidation, and the integration of additional pedagogically desirable features suggested by insights gained from pedagogic research studies.

Beyond tools development, corpus creation and corpus search, the pedagogic orientation of BACKBONE also includes the embedding of corpus-based learning materials and activities in e-learning and e-teaching contexts. The guiding principle is the integration of ‘focus on form’ activities within an overall communicative and collaborative learning environment. To support and combine these two complementary task and activity types, BACKBONE uses the authoring software [Telos Language Partner](#) (Kohn 2008) and the open source e-learning platform, [Moodle](#).

The pedagogical potential and added value of the BACKBONE corpora is explored and evaluated in 24 specially designed pilot courses. The range of settings includes CLIL-related foreign language in secondary, higher and vocational education, teaching English as a lingua franca (ELF), and community interpreter training in higher education. The courses cover all BACKBONE languages and corpora.

The project’s dissemination and exploitation strategy builds on open web access to the BACKBONE tools, the corpora and learning materials, “sand box” courses on the project’s Moodle site, and accompanying teacher workshops in the various partner countries. Close integration of all teacher training with real courses in regular school programmes is deemed essential and is given priority. Teacher training and support measures are reinforced by the general do-it-yourself quality of the BACKBONE approach and tools.

3. Preparatory pedagogic research

Pedagogic research in BACKBONE involves and combines three task dimensions: (a) an empirical “fields & needs” analysis to ensure a sound pedagogic contextualization of the project, (b) a state-of-the-art study about the theoretical-methodological foundations of a corpus-enhanced language learning & teaching approach, and (c) the formative and summative pedagogical evaluation of the BACKBONE tools, corpora and courses.

(a) “Fields & needs” analysis

The “fields & needs” analysis (Braun & Slater 2009) was employed to obtain an overview of those institutions willing to set up BACKBONE pilot courses and of the pedagogical environments in which the pilot courses would have to be implemented. A questionnaire approach was used to collect information about the different types of institutions, their technological infrastructure, and the teaching needs and approaches for which the pilot courses should be designed. The overall aim was to ensure that the corpora and the learning activities employed in the BACKBONE project were relevant for the teachers and learners participating in the pilot courses.

The picture that emerged from the completed questionnaires was that the regulations, circumstances and preferences of the pilot course institutions and teachers, and the needs and proficiencies of the learners were very diverse across the different institutions. Differences were found with regard to learning objectives, the ways in which the classes were structured and taught, and the length of time available for piloting. However, it also became evident that the course settings envisaged were all suitable for CLIL-related e-learning solutions and thus clearly within the target range of BACKBONE.

(b) State-of-the-art study: "Corpus-enhanced language learning & teaching"

The results of the state-of-the-art study are presented in two reports. The first one, "Spoken multimedia corpora for student-centred corpus exploration" (Braun 2009), takes a research perspective and looks at the theoretical and methodological foundations of the pedagogic application of corpora. Based on a review of recent research literature, the report argues for a pedagogic corpus approach that reaches beyond the mere pedagogic application of descriptively motivated corpora; it emphasizes the need for corpora to be designed from a specifically pedagogic vantage point right from the start. Using examples from the ELISA corpus, the report continues to illustrate how corpus-based work in the classroom can be expanded beyond the conventional methods of data-driven learning. In this way, it provides valuable orientation for corpus development in BACKBONE.

The second report, "Using pedagogic corpora for form and communication integrated learning" (Hoffstaedter, Kohn & Widmann 2009) specifies prototypical corpus-based learning tasks and activities that can serve as a model for the creation of language learning resources in BACKBONE. Topic oriented learning activities include listening and reading comprehension supported by multimedia learning modules, topic driven explorations of the corpora, as well as thematic internet explorations (e.g. WebQuests) using corpus material as an opener or starting point. Language oriented activities focus on vocabulary, grammar and communicative functions, but also on spoken discourse and regional and social varieties. It illustrates how the various modes and functions of the BACKBONE search tool can be used in Moodle for creating suitable exercises that combine a focus on form with a focus on communication.

4. Pedagogic corpus tool development

BACKBONE corpus tool development (3) covers corpus functions from transcription, annotation and enrichment to online search. Maintaining a consistent pedagogical orientation has been a constitutive principle of development. Since the BACKBONE interviews are intended for learning and teaching contexts, the BACKBONE **Transcriptor** (cf. Figure 1) (Pérez-Paredes, Alcaraz & Sánchez-Tornel 2011) employs an orthographical notation including punctuation; the punctuation conventions are slightly adapted to suit characteristics of spoken discourse. Pre-defined mark-up codes are used to specify breaks, truncations, alternatives, comments, etc. Fillers, repetitions, and hesitation phenomena are accounted for if considered to be communicatively relevant. Interview transcripts are divided into thematic sections; a time-stamping function is used to synchronize these sections with their corresponding video/sound files.

The BACKBONE **Annotator** (cf. Figure 1) (Pérez-Paredes & Alcaraz 2009; Pérez-Paredes, Alcaraz & Sánchez-Tornel 2011) operates on interviews and short transcript sections, and produces a corpus XML file. The interview sections are annotated in a drag & drop fashion with categories deemed relevant by the annotator-teacher, e.g. thematic, grammatical, lexical, and textual categories, and CEFR level specification. Words or phrases in a section that fit a certain category can be marked. The aim of annotation is not a classificatory one. The categories are rather meant to support meaningful searches; they can be defined by the annotator/teacher and thus tailored to capture the pedagogic potential of each individual corpus.

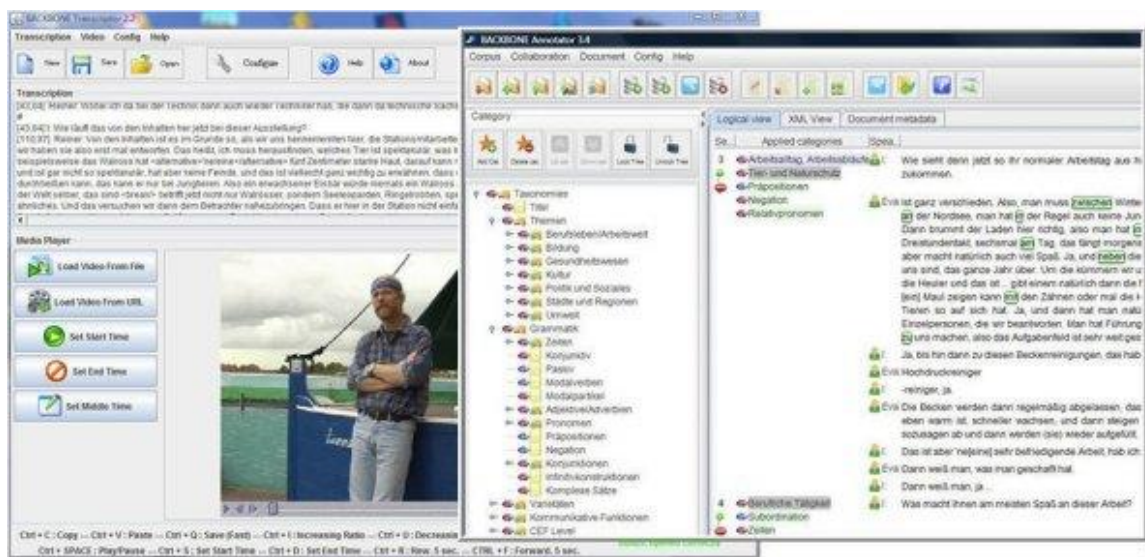


Figure 1. BACKBONE Transcriber and Annotator.

The Annotator can be used in collaboration mode. In this case, a web service links it to the online BACKBONE **Corpus Management Tool** (CMT), which supports simultaneous annotation, i.e. annotation by different annotators of different interviews of the same corpus at the same time (Pérez-Paredes, Alcaraz & Sánchez-Tornel 2011). In addition, enrichment resources in the form of ready-made learning modules and instructions for communicative and collaborative learning exploration are managed via the BACKBONE **Virtual Resource Pool** (VRP) and linked to interview sections (Kohn, Widmann, Wetzel & Hoffstaedter 2011). The resources themselves are stored in the VRP in a virtual fashion, i.e. together with their web address and a short description. Teachers can browse the listed resources, select the ones they need and drop them into a Resource Sheet (comparable to a shopping cart). During the annotation procedure, they can create links from interview sections to relevant Resource Sheets.

The annotated and enriched interviews are accessed by teachers and learners via the online BACKBONE **Corpus Search Site**, which has been designed to operate on a pedagogically annotated XML corpus file created with the Transcriber/Annotator, enriched with links to learning resources in the VRP, and stored in the Corpus Management Tool (Kohn, Widmann & Wetzel 2011). BACKBONE Search offers five search modes: 'Browse', 'Section search', 'Concordances', 'Co-occurrence', and 'Lexical lists'.

'Browse' (cf. Figure 2) displays all interviews contained in a corpus along with short descriptions. Entire interviews can be viewed, listened to and read, which facilitates contextualization and discourse authentication. In addition, the interview audio can be downloaded for further use, for instance, on an MP3 player.

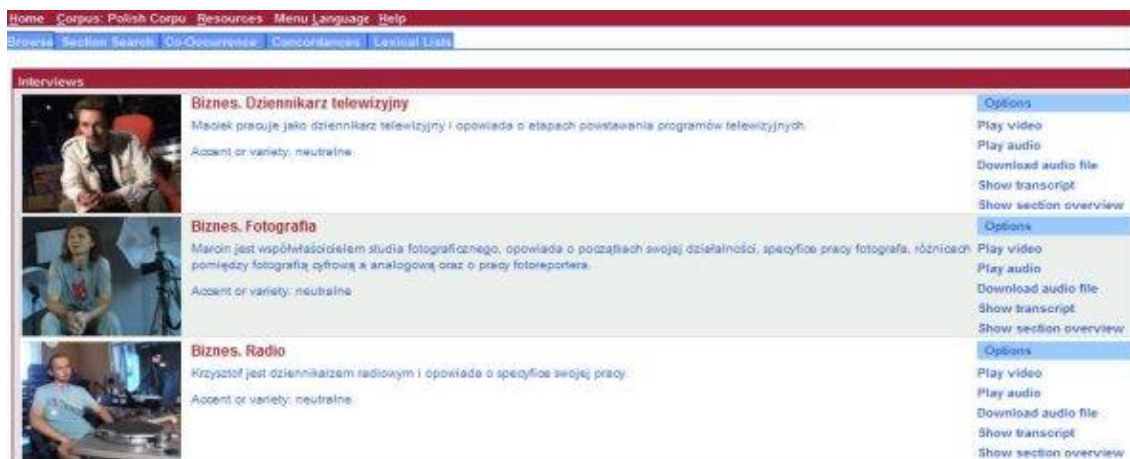


Figure 2. BACKBONE Search – 'Browse'.

It is also possible to use 'Section overview' to see all the thematic sections which make up an interview. Each section is displayed by title along with information regarding duration and length; links provide access to the corresponding video and sound files and to the annotated section itself.

'Section search' presents the annotation/search category tree (cf. Figure 3) and is used to search for individual interview sections that comply with a specified combination of thematic and linguistic annotation categories.

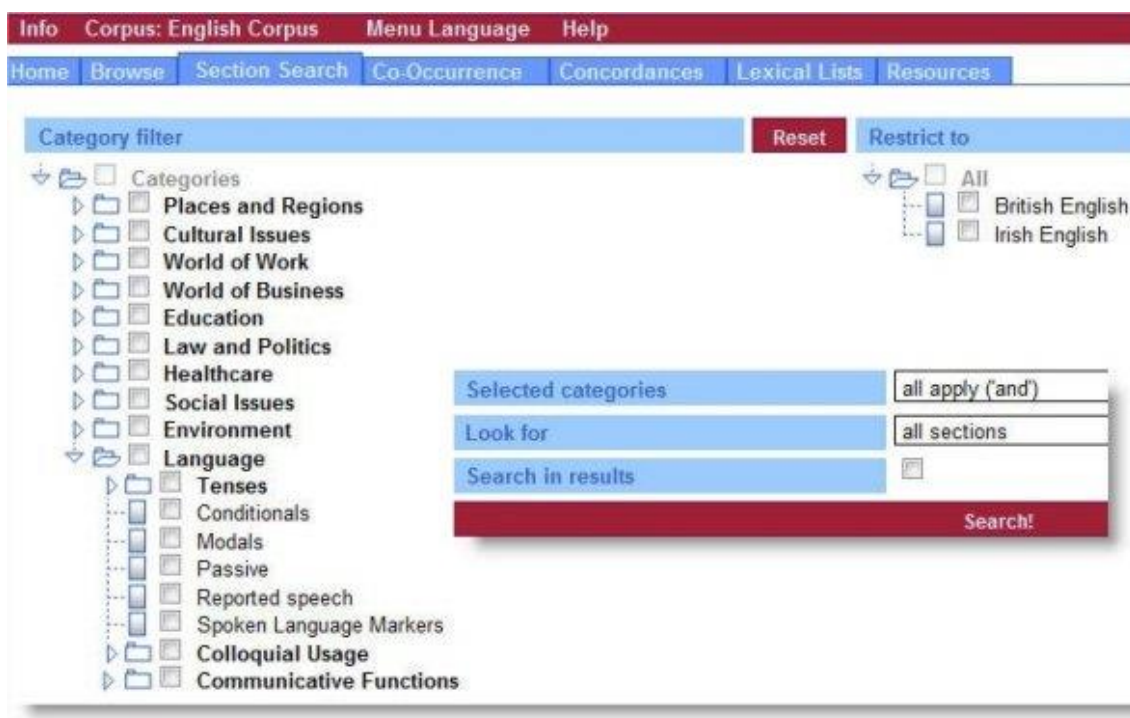


Figure 3. BACKBONE Search – 'Section search': annotation/search category tree.

In a section found as a search result (cf. Figure 4), all words and phrases that have been marked (during annotation) as 'satisfying' the respective category can now be highlighted. It is also possible to access the corresponding video and sound files and to call up the entire interview to see the section in a wider context.



Figure 4. BACKBONE Search – ‘Section search’: results.

Other specifications can be used in combination with selected annotation/search categories to further restrict and focus a search. This concerns, in particular, narrowing a search to selected sub-corpora (e.g. British or Irish English) or to sections that have been enriched with an attached ‘Resource sheet’ containing additional learning resources from the VRP.

‘Co-occurrence’ (cf. Figure 5) lists sections that contain a number of specified words in free distribution. Two wildcards replacing any number of characters (*) or one single character (?) can be used to include morphological word families.

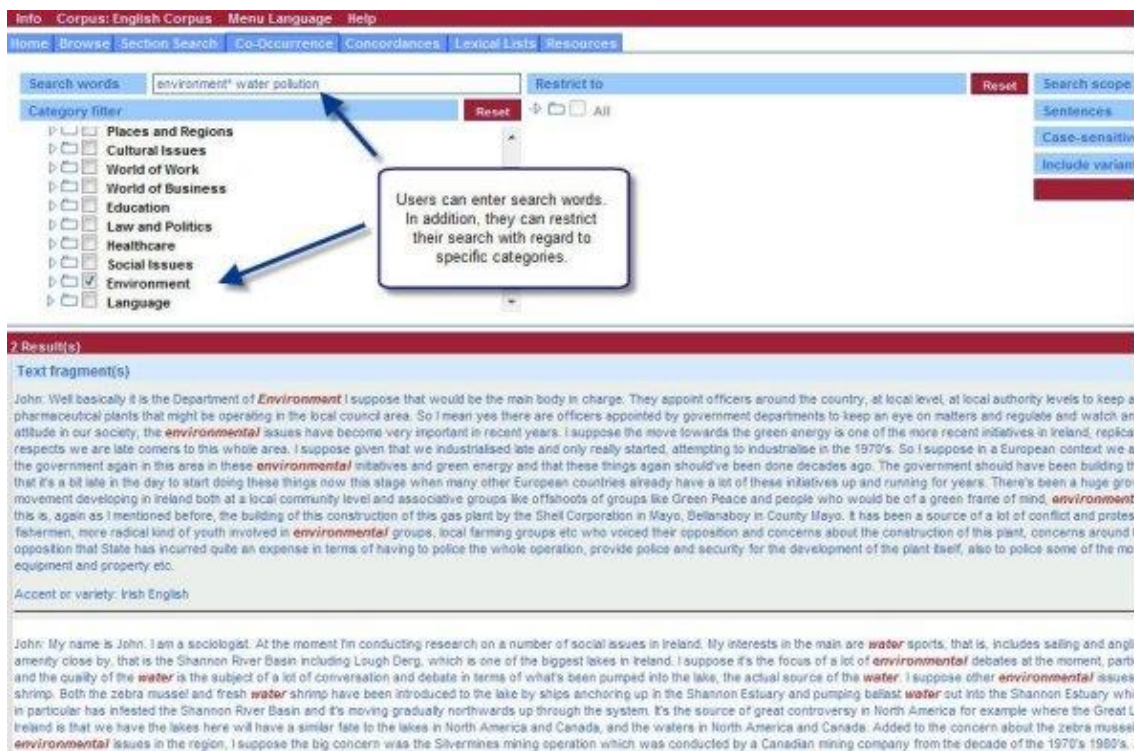


Figure 5. BACKBONE Search – ‘Co-occurrence’.

'Concordances' (cf. Figure 6) produces lines of text with keywords in context as with other KWIC concordancers. Three words can be used along with the two wildcards '*' and '?'.

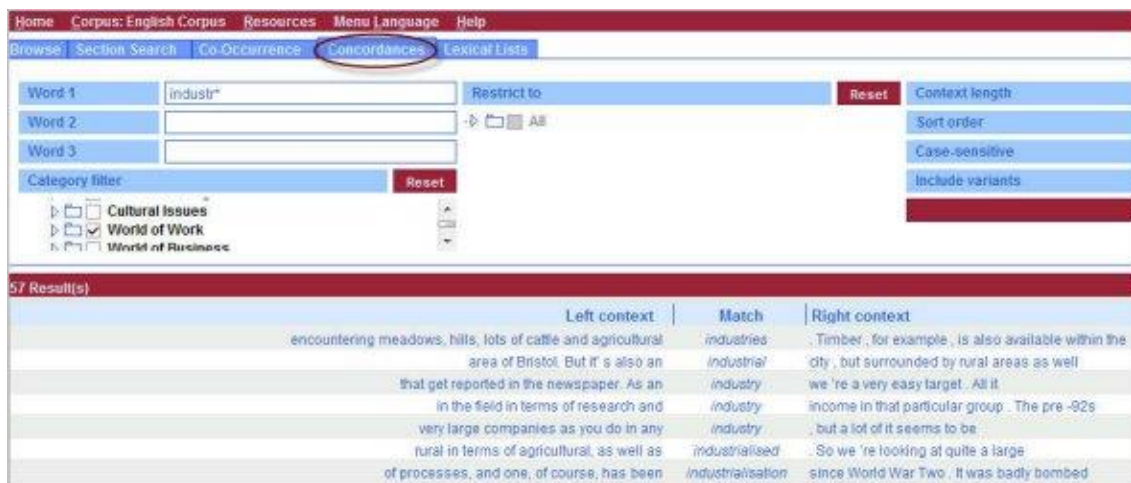


Figure 6. BACKBONE Search – 'Concordances'.

Both co-occurrence and concordance searches can be combined with selected annotation/search categories thereby limiting the search scope to sections that deal with a certain topic, exhibit certain grammatical properties, or belong to a preferred CEFR level. Restriction to certain language varieties is possible as well.

'Lexical lists' enables users to display either all occurring words or all words and phrases that have been marked by a certain annotation category (cf. Figure 7). The options include 'all words' and 'annotated words and phrases'. Both lists also indicate frequency of occurrence and provide access to concordances; they can be based on the entire corpus or only on those interview sections that fit a certain annotation category or category combination.

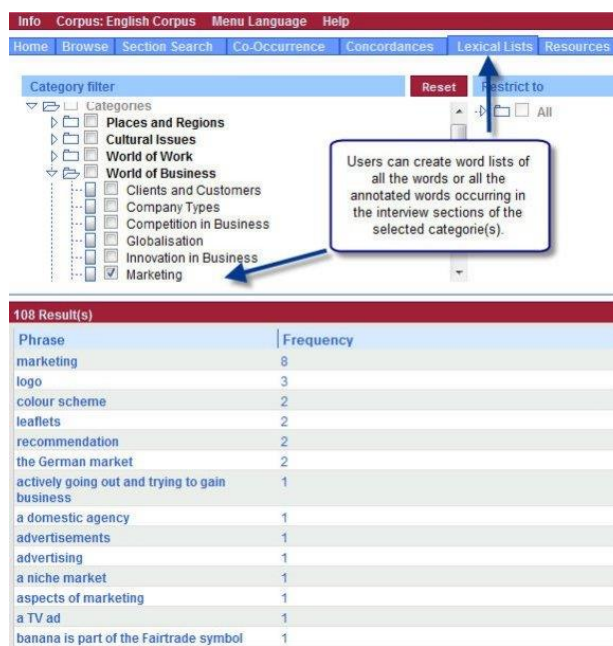


Figure 7. BACKBONE Search – 'Lexical lists'.

As in the other search modes, a preferred thematic focus can be specified by selecting the appropriate annotation/search or variety category. While substantial parts of these corpus tools were available from the SACODEYL project (see chap. 1, above), additional major extensions were designed and implemented in BACKBONE. Beyond necessary consolidation work, these extensions include significant improvements in functionality, in particular with regard to new mark-up and time-stamping features in the transcription procedure, collaborative and simultaneous corpus annotation and corpus management, multi-layered annotation, a web service integrating enrichment resources both during annotation and search, alignment of formats for video streaming and download of sound files, as well as lexical pattern search.

5. Pedagogic corpus creation

The BACKBONE suite of pedagogic corpora consists of **seven sub-corpora** of video-recorded interviews in English, French, German, Polish, Spanish and Turkish, as well as in European manifestations of English as a Lingua Franca (ELF) (Hoffstaedter 2011). English is covered by 50 interviews, including 25 British and 25 Irish interviews; all the other languages are represented with 25 interviews each. The ELF corpus contains a total of 50 English interviews as well, with 10 native speakers from each of the five base languages French, German, Polish, Spanish and Turkish. The interviews are an average length of 10 minutes.

The BACKBONE corpus compilation procedure includes three main tasks: interview collection, transcription and annotation, and development and embedding of pedagogic enrichment resources. The **first compilation step**, interview collection, involves in particular the identification of pedagogically relevant topics, the specification of desirable speaker characteristics, recruiting of interviewees, as well as the organization and video-recording of the interviews. The topics covered in the BACKBONE corpora have been identified in collaboration with teachers and through the analysis of relevant course books. The thematic areas chosen emphasize a regional perspective and include culture, world of work, urban and rural life, social issues, health and social security, education, environment, government and politics (Table 1).

Cultural issues	customs/traditions, food, special days, ceremonies
	arts (music, movies, youth culture)
	sports
	new technologies
Economy	fishing, automotive industry and other industries
	agriculture
	globalisation
World of Work	occupations, working conditions, trade unions
	work placement, internship
Urban and rural life	living in a city or in a mega-city (London)
	suburban life
	rural life
Social issues	minorities and fringe groups

	ethnic groups
	immigration
	multicultural society
Health & social security	national health system and hospitals
	health professions
	welfare and social benefits
Education	educational system and institutions
	vocational education
	educational mobility
The environment	climate change
	traffic and pollution
	renewable energies (e.g. solar energy, wind farms)
	environmental policy in industrial companies
Government & politics	political system, institutions and parties
	mayor, city council, local government
	Europe

Table 1: Thematic areas covered in the BACKBONE corpora.

Across these thematic areas, each corpus tries to strike a balance between the range of coverage and the thematic preferences set by the envisaged BACKBONE pilot courses. The following summary account gives an idea of the breadth of topics dealt with across the interviews in the various corpora.

The **British English interviews** (4) were recorded in selected regions in the UK including the counties of Surrey (Guildford), Somerset (Bristol, Cleve, Martock and Taunton), Devon (Plymouth) and the West Midlands (Birmingham). Some interviews focus on other parts of the country (Cheshire, Derbyshire, Lancashire). The interviewees include the Managing Director of a science park, a senior nurse at a special care baby unit, a virtual learning specialist, a wedding planner, and a lawyer.

The **Irish English interviews** (5) were recorded in selected regions in Ireland, in particular the counties of Cork, Tipperary, Kerry, Dublin, Laois, Roscommon, Clare and Mayo. The participants include teachers from primary, secondary and tertiary education, a community dietician, a taxi driver, a jewellery shop assistant, various sports players and enthusiasts, a Product Development Officer for Failte Ireland, and a farmer.

Although most of the **French interviews** (6) were recorded in the Jura region, the people selected for the interviews came from different parts of France or French speaking areas: Jura, Bresse, Paris, Lorraine, Provence, and Africa. The interviewees include trainers in a business school, students of Medicine, Politics, and Business Administration, a salesman, a retired Post Office inspector, a chief accountant, a webmaster, an IT engineer, and the President of a Football Association.



The **German interviews** (7) were recorded in different regions and cities including Southern Germany (Lake Constance, Baden, Swabia), the Rhine-Ruhr area, Northern Germany (Dithmarschen/Schleswig-Holstein), and Berlin; one interview was recorded in Austria (Vienna). The interviewees include teachers, a city councillor, Green Party activists, a general practitioner, automotive engineers, a tourism director, a delicatessen shop owner, a fisherman, workers in nature conservation and animal-welfare, an artist, and a writer.

The **Polish interviews** (8) were recorded in central Poland with speakers of the standard language, including the owner of a computer shop, an interpreter, a staff member from the Polish Agency for Enterprise Development, a television journalist, a make-up artist, as well as people discussing social topics such as sexual minority, icons of pop culture or extraordinary hobbies.

The **Spanish interviews** (9) were recorded with speakers from different regions: Aragon (Zaragoza), Andalusia (Jaen, Seville, Granada), Cantabria (Celis/Santander), Castilla-La Mancha (Albacete), Comunidad Valenciana (Alicante), Galicia (La Coruña), and Murcia. The interviewees include a top researcher, an ex-lawyer, teachers, an NGO volunteer, a doctor, a librarian, a sportswoman, a former quality control laboratory worker, a folklorist, two young entrepreneurs, a telemarketer, a bio farmer, and a clerk.

The **Turkish interviews** (10) were recorded in Kayseri, Central Anatolia, and involve speakers of standard Turkish born and raised in different regions in Turkey, including a housewife, an insurance expert, computer specialists, a dentist, a librarian, mechanical engineers, a science teacher, a pharmacist, two architects, a florist, a catering manager, a shop owner, a medical doctor, a lawyer, a banker, an optician, a project administrator, and a hairdresser.

The **English as a Lingua Franca (ELF) interviews** (11) were recorded with non-native speakers of English in France, Germany, Poland, Spain and Turkey. The interviewees are all used to speaking English in their work environments or privately on a regular basis; and they come from a wide range of different professional backgrounds. The topics they speak about are similar to those of the native speaker interviews

In a **second compilation step**, the recorded interviews were transcribed and time-stamped with the BACKBONE Transcriptor; the transcribed video recordings were analysed and annotated with the BACKBONE Annotator with regard to pedagogically relevant characteristics including e.g. topic, grammar, communicative functions and CEFR level.

The **third compilation step** concerned supplementing each corpus with pedagogic enrichment resources. Besides the interview video/sound files, this concerns in particular two types of language learning resources: (a) ready-made learning modules for self-study, and (b) task suggestions and instructions for exploratory and communicative learning activities combining web tools such as forums, chats or wikis with corpus explorations and classroom interactions. The development of the BACKBONE language learning resources was informed by the analysis of language learning tasks and activities undertaken in Hoffstaedter, Kohn & Widmann 2009. The created resources were stored in the Virtual Resource Pool (VRP) and linked to the interview sections as part of the annotation process.

The **ready-made learning modules** consist of web-based multimedia exercises created with the authoring software Telos Language Partner. The exercises focus on combinations of listening comprehension, vocabulary practice and grammar learning in a variety of task formats including multiple choice, true/false, select, gap filling, or drag & drop (cf. Figure 8).

Figure 8. Telos learning module.

The **exploratory and communicative learning activities** consist of suggestions and instructions for teachers who want to use the BACKBONE corpora in an e-learning environment. Relevant web tools include forums, chats and wikis or the Moodle glossary for the collaborative creation of topic-specific dictionaries. Exploratory and communicative activities have been developed for each of the main topics represented in the corpora and they are available as PDF files, each covering a number of activity sheets on different aspects of the respective topic. A typical activity sheet contains task descriptions and suggestions concerning a suitable interaction mode (e.g. individual work, pair, or group work), a specification of suitable web tools to be used, and links to useful websites. Typical learning activities include topic-driven corpus explorations requiring learners to study certain interview sections with regard to specific questions, as well as topic and task-driven exploratory internet research using the corpus materials as an opener or starting point but reaching out beyond the immediate scope of the corpus.

The exploratory learning activities may also include vocabulary explorations using specific search options and features of BACKBONE Search such as Co-occurrence, Concordance or Lexical lists.

The language learning resources can be accessed in Backbone Search either via the menu tab 'Resources' or in 'Browse' and 'Section search' via the interview sections to which they have been attached. Additional information on both the video interviews and the language learning resources is available on the BACKBONE project website via the menu tabs 'Corpora & search' and 'Project documentation'.

6. Pilot courses and evaluation

The BACKBONE approach was explored and evaluated in 24 **pilot courses** (Kohn & Hoffstaedter 2011) comprising 9 foreign language courses in secondary education, 8 in higher education and 4 in vocational education, as well as 3 multilingual community interpreter courses in higher education. The courses cover the 6 target languages: English, French, German, Polish, Spanish and Turkish. In terms of pedagogic approach, they all combine a CLIL orientation with e-learning support in a blended learning environment.

Topics include e.g. 'Multicultural society', 'Economic globalization', 'The Irish sports camogie and hurling', 'Francophonie', 'La vie d'une étudiante en médecine', 'The Berlin Wall', 'Health issues', 'Media and music in Poland', 'Amphibious vehicles', 'Certification and safety in the car industry', 'Mobility and the oil industry', as well as various types of 'Education'. Three of the higher education courses are embedded in language teacher study programmes, one of which specifically addresses the needs of teacher students with regard to learning to teach English as a lingua franca (ELF).

A characteristic foreign language course unit may, for example, start off with video-based awareness raising and forum discussion activities, followed by comprehension checks with ready-made TELOS modules; it may then continue with collaborative thematic and/or linguistic corpus explorations, spoken interaction in class, or summary writing tasks in a forum or as an individual assignment. For more information, see the demo courses on the BACKBONE website under the menu tab "Courses".

The multilingual community interpreter courses follow a somewhat different pedagogical approach specifically adapted to the requirements of interpreter training. They typically involve up to six languages and implement corpus-based activities focusing on interpretation-related skills such as active listening, memory training, anticipation, note taking, and consecutive production.

To facilitate pilot course implementation, a **teacher training and support area** was set up on the BACKBONE website. It provides information, instructions and hands-on practice regarding the BACKBONE pedagogic approach, the BACKBONE search tool and corpora, creation of corpus-based learning units, use of Moodle for designing online courses, as well as do-it-yourself tools for developing one's own customized corpora and learning materials. This module is supplemented by demo courses and a course template for e-learning sequences. Furthermore, teacher training workshops were set up in connection with project dissemination and exploitation activities to provide opportunities for networking and collaboration. The pilot course area under the menu heading 'Courses' is kept available for hosting "guest courses" from external cooperation partners.

The pilot courses made it once again clear that e-learning – in particular with regard to languages – is still a rather marginal phenomenon and one of the major challenges for teacher training and support. Preparatory field analyses carried out in BACKBONE (Braun & Slater 2009) as well as survey studies from the Comenius network projects EcoMedia and Wide Minds (Kohn & Esteves 2009; Kohn, Glombitza & Helbich 2008) provide strong evidence that for many language teachers and course providers in secondary, higher and vocational education a smooth and seamless pedagogic integration of e-learning is not yet a practical reality. Throughout the BACKBONE pilot course activities, it was thus necessary for local course supervisors and e-learning experts from the BACKBONE team to provide intensive and continuous teacher support. In several cases, problems concerning local technological infrastructure – in particular regarding availability of computer rooms, restricted web access due to security settings, or limited technical support at schools – persisted despite previous checks and required continuous monitoring and "tailored" solutions.

These technological problems and challenges were also emphasized in the **pedagogical assessment and evaluation** of the BACKBONE approach based on questionnaire,

interview and performance feedback from all 24 pilot courses spread across four educational settings from secondary, higher and vocational education to interpreter training (Braun & Slater 2011). Despite all technological flaws and misgivings, however, the BACKBONE approach received an overall positive reception. The topics covered in the video interviews were judged highly relevant by the teachers and pedagogically suitable for initiating a wide range of learning activities integrating content and language. Students were engaged in thematic vocabulary explorations and were encouraged to talk and write. The diversity of voices and regional accents significantly helped them develop their aural and comprehension skills. In addition, the English as a lingua franca corpus provided innovative opportunities for improving awareness and comprehension in relation to a wide variety of European non-native speaker manifestations of English. While the BACKBONE topics and interviews are reported to fit in well and also seem to be easy to integrate, it should be emphasized that the thematic requirements and preferences across courses, course types and educational settings outside the initial piloting scenarios will certainly be more varied. In this connection, the need for American English and South American Spanish was specifically mentioned. There is thus clearly a place for the do-it-yourself quality of the BACKBONE tools, which allow for customizing small corpora to given course curricula and course contents – ideally in cooperation with course content providers (e.g. publishers).

The learning resources provided in connection with selected interviews and interview sections received very positive ratings and comments. The Telos Language Partner modules were valued for their focus on form in thematic contexts, immediate feedback and scoring options and their self-study potential, in particular for weaker students; in the case of the exploratory and communicative activities, it was emphasized that they save preparation time and provide useful suggestions for collaborative interaction. Evaluation of the BACKBONE search tool was obviously influenced by the users' degree of familiarity with computers and e-learning: some found it easy to use; others felt it was functionally and pedagogically difficult to understand. Several teachers emphasized the need to compensate for the limited content and language range of a small corpus. They suggested using the search tool for checking out self-study exploration paths that lead to meaningful search results and thus avoid unnecessary dissatisfaction and frustration on the students' part. Working with Moodle was generally given high positive ratings; some teachers emphasized its potential for easy and cohesive course development and support, collaborative learning with forums and wikis, as well as online writing assignments with individual feedback. At the same time, however, the need for better design transparency and more detailed instructions was mentioned. In some cases switching between Moodle and the search interface was felt to be somewhat confusing. It was also pointed out that careful lesson planning can initially be quite time-consuming.

All in all, though, integrating BACKBONE corpora within a Moodle-based e-learning environment combining form and content-focused self-study with collaborative communication and interaction seems to be well suited to support content and language integrated learning (CLIL); and, in a somewhat different vein, it also helps to cater to the challenges of teaching English with an emphasis on lingua franca communication. The foreign language students' perceived learning success was encouragingly high; and a majority voiced their interest in continuing this teaching/learning approach. This finding was matched by the overall positive learning success rating obtained from teachers – albeit with repeated reference to the need for a sound pedagogical embedding. In the case of consecutive interpreting, the BACKBONE resources proved highly relevant for step-by-step interpreter training both for classroom and self-study scenarios. The interpreting-related activities clearly helped students improve their interpreting skills; they were particularly useful in bridging undergraduate to postgraduate levels.

It can be concluded that e-learning activities, used in the right way, significantly foster key principles of communicative and constructivist language learning, in particular the principles of authentication, learner autonomy and collaboration. In this connection it should be added, however, that the computer lab does not provide the most suitable

conditions for tapping the language learning potential offered by e-learning. Using a laptop in conjunction with data projection and internet access in class and individual and/or collaborative self-study outside class as part of homework or independent study may prove to be more efficient, albeit more difficult to implement and supervise pedagogically.

These clearly positive piloting results must not eclipse the negative comments, which were mostly concerned with IT equipment and interface functionality, course navigation and pedagogical issues. However, these negative evaluation results are put into perspective by the fact that for most institutions, teachers and students involved in the pilots the integration of e-learning activities was an entirely new experience. Shortcomings due to insufficient technological infrastructure and lack of technological and/or pedagogical familiarity and expertise on the part of the users are thus hardly surprising. Instead of disproving the pedagogical potential and feasibility of e-learning, the recorded weaknesses rather underline the urgent need for

- raising awareness regarding possibilities, limitations and challenges,
- ensuring adequate technological support both in terms of equipment and staff,
- implementing continuous and focused teacher training measures.

However, these measures may not be sufficient due to the 'classroom approach' firmly established in educational institutions from schools and universities to adult education and vocational training. The pedagogic potential of e-learning for supporting autonomous, authenticated and collaborative learning outside classroom hours can only be fully exploited if learning activities are supported by teachers who are available as need arises. This requires teachers to be more flexible than is usually possible within the traditional workload management system based on face-to face teaching hours.

References

- Aston, G. (2002). The learner as corpus designer. In B. Kettemann & G. Marko (eds.). *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, 9-25.
- Aston, G., Bernadini, S. & Stewart, D. (2005). *Corpora and Language Learners*. Amsterdam: John Benjamins.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (eds.) (1999). *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.
- Boulton, A. (2011). Data-driven learning: The perpetual enigma. In S. Goźdz-Roszkowski (ed.). *Explorations across Languages and Corpora*. Frankfurt/M: Peter Lang, 563-580.
- Braun, S. & Slater C. (2009). Pedagogical fields and needs analysis. BACKBONE Report 3.2 [<http://projects.ael.uni-tuebingen.de/backbone> > Project documentation].
- Braun, S. & Slater C. (2011). Pedagogical assessment and evaluation. BACKBONE Report 3.3 [<http://projects.ael.uni-tuebingen.de/backbone> > Project documentation].
- Braun, S. (2005). From pedagogically relevant corpora to authentic language learning contents. *ReCALL*, 17(1), 47-64.
- Braun, S. (2007). Integrating corpus work into secondary education: From data-driven learning to needs-driven corpora. *ReCALL*, 19(3), 307-328.
- Braun, S. (2010). Getting past 'Groundhog Day': Spoken multimedia corpora for student-centred corpus exploration. In T. Harris & M. Moreno Jaén (eds). *Corpus Linguistics in Language Teaching*. Frankfurt: Peter Lang, 75-98.

- Braun, S., Kohn, K. & Mukherjee, J. (2006). *Corpus Technology and Language Pedagogy: New Resources, New Tools, New Methods*. Frankfurt/M: Peter Lang.
- Flowerdew, L. (2009). Applying corpus linguistics to pedagogy. A critical evaluation. *International Journal of Corpus Linguistics*, 14(3), 393-417.
- Granger, S., Dagneaux, E. & Meunier, F. (2002). *The International Corpus of Learner English*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Guth, S. & Helm, F. (eds.) (2010). *Telecollaboration 2.0*. Bern: Peter Lang.
- Hoffstaedter, P. & K. Kohn (2009). Real language and relevant language learning activities: insights from the SACODEYL project. In A. Kirchhofer & J. Schwarzkopf (Eds.), *The Workings of the Anglosphere. Contributions to the Study of British and US-American Cultures*. Trier: WVT, 291-303.
- Hoffstaedter, P. (2011). Compilation of the BACKONE pedagogical corpora. BACKBONE Report 5.1-7 [<http://projects.ael.uni-tuebingen.de/backbone> > Project documentation].
- Hoffstaedter, P., Kohn, K. & Widmann, J. (2009). Corpus-enhanced language learning and teaching. Part 2: Using pedagogic corpora for form and communication integrated learning. BACKBONE Report 3.1 (part b) [<http://projects.ael.uni-tuebingen.de/backbone> > Project documentation].
- Johns, T. & King, Ph. (1991). *Classroom Concordancing*. Birmingham: University of Birmingham.
- Kohn, K. & Hoffstaedter, P. (2011). Pilot courses. BACKBONE Report 6 [<http://projects.ael.uni-tuebingen.de/backbone> > Project documentation]
- Kohn, K. & Warth, C. (2011). *Web Collaboration for Intercultural Language Learning. A Guide for Language Teachers, Teacher Educators and Student Teachers. Insights from the icEurope Project*. Münster: MV-Wissenschaft.
- Kohn, K. (2008). Telos Language Partner: DIY authoring for content-based language learning. In A. Gimeno (ed.). *Computer Assisted Language Learning: Authoring Tools for Web-Based CALL*. Valencia: Universidad Politécnica de Valencia, 157-174.
- Kohn, K. (2009). Computer assisted foreign language learning. In K. Knapp & B. Seidlhofer (eds.). *Foreign Language Communication and Learning. Handbooks of Applied Linguistics*, Volume 6. Berlin: Mouton-de Gruyter, 573-603.
- Kohn, K., Widmann, J. & Wetzels, D. (2011). Pedagogic corpus search tool. BACKBONE Report 4.3 [<http://projects.ael.uni-tuebingen.de/backbone> > Project documentation].
- Kohn, K., Widmann, J., Wetzels, D. & Hoffstaedter, P. (2011). Pedagogic corpus enrichment tools. BACKBONE Report 4.2 [<http://projects.ael.uni-tuebingen.de/backbone> > Project documentation].
- Kohn, K. & Esteves, M. (2009). The use of educational e-learning equipment and applications in foreign language classes: a questionnaire-based survey. Wide Minds Comenius Network Report (Workpackage 5: Developing Multilingualism through Digital Content). [<http://projects.ael.uni-tuebingen.de/backbone> > Project documentation].
- Kohn, K., Glombitza, A. & Helbich, G. (2008). Perceived potential of educational ICT in European schools. EcoMedia Socrates Comenius 3 Network (ed.). *ICT in European Schools, ePortfolios and Open Content*. St. Michael: der wolf verlag, 35-42. [<http://projects.ael.uni-tuebingen.de/backbone> > Project documentation].

Kohn, Kurt, Petra Hoffstaedter & Johannes Widmann (2010). BACKBONE – pedagogic corpora for content & language integrated learning. In A. Gimeno Sanz (ed.). *New Trends in Computer-Assisted Learning: Working Together*. Madrid: Macmillan ELT, 157-162.

Pérez-Paredes, P. & Alcaraz-Calero, J.M. (2009). Developing annotation solutions for online data driven learning. *ReCALL*, 21(1), 55-75.

Pérez-Paredes, P. (2010). Corpus linguistics and language education in perspective: appropriation and the possibilities scenario. In T. Harris & M. Moreno Jaén (eds). *Corpus Linguistics in Language Teaching*. Frankfurt: Peter Lang, 53-73.

Pérez-Paredes, P., Alcaraz, J.M. & Sánchez-Tornel, M. (2011). Pedagogic corpus compilation. BACKBONE Report 4.1 [<http://projects.ael.uni-tuebingen.de/backbone> > Project documentation].

Sinclair, J. (ed.) (1987). *Collins COBUILD English Language Dictionary*. London: Collins.

Sinclair, J. (ed.) (1990). *Collins COBUILD English Grammar. Helping Learners with Real English*. London: Collins.

Sinclair, J. (ed.) (2004). *How to use Corpora in Language Teaching*. Amsterdam: John Benjamins.

Swain, M. (2005). The Output Hypothesis: Theory and Research. In E. Hinkel (2005). *Handbook on Research in Second Language Teaching and Learning*. Mahwah, NJ: Lawrence Erlbaum, 471-484.

Tribble, Ch. & Jones, G. (1990) *Concordances in the Classroom: A Resource Book for Teachers*. Harlow: Longman.

Tribble, Ch. (1997). Improvising corpora for ELT: quick-and-dirty ways of developing corpora for language teaching. In B. J. Lewandowska-Tomaszczyk & P. J. Melia (eds.). *PALC-97: Practical Applications in Language Corpora*. Łódź: Łódź University Press, 106-117.

Wichman, A., Fligelstone, S., McEnery, T. & Knowles, G. (eds.) (1998). *Teaching and Language Corpora*. London: Longman.

Widdowson, H. G. (1991). The description and prescription of language". In J. Alatis (ed.). *Linguistics and Language Pedagogy: the State of the Art*. Washington, DC: Georgetown University, 11-24.

Widdowson, H. G. (1998). Context, community and authentic language. *TESOL Quarterly*, 32(4), 705–716.

Widdowson, H.G. (2003). *Defining Issues in English Language Teaching*. Oxford: OUP.

Widmann, J., Kohn, K. & Ziai, R. (2010). The SACODEYL search tool: exploiting corpora for language learning purposes. In Frankenberg-Garcia, A., Flowerdew, L. & Aston, G. (eds.). *New Trends in Teaching and Language Corpora. Proceedings of the TaLC 2008*. London: Continuum, 321-327.

Websites

ELISA corpus: <http://projects.ael.uni-tuebingen.de/elisa> (Last accessed 17 Sept. 2012)

SACODEYL project: <http://www.um.es/sacodeyl> (Last accessed 17 Sept. 2012)

BACKBONE project: <http://projects.ael.uni-tuebingen.de/backbone> (Last accessed 17 Sept. 2012)

BACKBONE corpus search: <http://purl.org/backbone/searchtool> (Last accessed 17 Sept. 2012)

Notes

[1] The BACKBONE project has been funded with support from the European Commission. This publication reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

[2] In this connection also see Swain's (2005) observation concerning the asymmetry between reception and production in immersion classes and her call for more (and "pushed") output processing.

[3] For more information and download options, please see the BACKBONE website. Development of the BACKBONE tools was carried out by two teams: (a) Transcriptor, Annotator and Corpus Management Tool (Universidad de Murcia: Pascual Pérez-Paredes, Jose María Alcaraz Calero, María Sánchez Tornel) and (b) Virtual Resource Pool (VRP) and Search Site (University of Tübingen: Kurt Kohn, Johannes Widmann, Dominikus Wetzel).

[4] Corpus compilation and learning materials: Sabine Braun and Catherine Slater (University of Surrey).

[5] Corpus compilation: Fiona Farr, Michelle O'Shea and Valerie Rabbette (University of Limerick); learning materials: Elaine Riordan (University of Limerick).

[6] Corpus compilation: Michel Meuret, Denis Trossat and Bernard Vallauri (Chambre de Commerce du Jura); learning materials: Michel Meuret and Bernard Vallauri (Chambre de Commerce du Jura).

[7] Corpus compilation and learning material: Petra Hoffstaedter (STZ Sprachlernmedien, DE).

[8] Corpus compilation: Jakub Karykowski (Academy of Humanities and Economics Łódź, PL); learning material: Sebastian Ostalak (Academy of Humanities and Economics Łódź, PL).

[9] Corpus compilation and learning materials: María Sánchez Tornel and Pascual Pérez-Paredes (Universidad de Murcia, ES).

[10] Corpus compilation: Doğan Bulut (Melikşah University, TR), Gullu Cinek, Ömer Erdoğan, Sevgi Erel, Ozlem Altun (Erciyes University, TR); learning materials: Doğan Bulut (Melikşah University, TR), Ömer Erdoğan and Sevgi Erel (Erciyes University, TR).

[11] Corpus compilation: Bernard Vallauri and Michel Meuret (Chambre de Commerce du Jura, FR), Johannes Widmann (University of Tübingen, DE), Jakub Karykowski (Academy of Humanities and Economics Łódź, PL), María Sánchez Tornel and Pascual Pérez-Paredes (Universidad de Murcia, ES), Doğan Bulut (Melikşah University, TR), Sevgi Erel and Ömer Erdoğan (Erciyes University, TR); learning materials: Johannes Widmann, Philipp Glaser and Matthias Kobsa (University of Tübingen, DE).

[Top](#)