



EUROCALL 2009 Conference

New trends in CALL: Working together

Gandia Campus

Universidad Politécnica de Valencia (Spain)

9-12 September 2009



UNIVERSIDAD
POLITECNICA
DE VALENCIA

BACKBONE – Pedagogic Corpora for Content & Language Integrated Learning

Kurt Kohn^{a,*}, Petra Hoffstaedter^b, Johannes Widmann^a

^a*University of Tübingen, Applied English Linguistics, Wilhelmstr. 50, D-72074 Tübingen, Germany*

^b*Petra Hoffstaedter, STZ Sprachlernmedien, Rheingasse 10, D-78462 Konstanz, Germany*

Abstract

BACKBONE is a European LLP (Languages) project involving 8 partners from 7 European countries: France, Germany, Ireland, Poland, Spain, Turkey and UK. The project addresses language learning needs with regard to pedagogically neglected languages and varieties including lesser taught languages, regional & socio-cultural varieties of more frequently taught languages, and European non-native speaker manifestations of English as a lingua franca. It proposes a pedagogically motivated corpus approach that enables teachers to record, annotate, and enrich spoken discourse resources. Based on the principles of learner and teacher autonomy, authenticated learning, and collaboration, the BACKBONE approach supports collaborative pedagogic deployment of CLIL-related corpus resources in Moodle courses. BACKBONE develops pedagogically motivated tools for corpus management, annotation, enrichment, and search. Seven sub-corpora of video-recorded interviews are compiled and enriched with learning materials: English, French, German, Polish, Spanish and Turkish as well as English as a lingua franca. Pedagogic piloting is carried out in CLIL settings in secondary, higher and vocational education.

© 2010 EUROCALL All rights reserved

Keywords: Pedagogic corpora; CLIL; second language learning; foreign language learning; e-learning; moodle

1. Introduction & contextualisation

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

* Contact author: Tel.: +49 7071 297 2377; Fax: +49 7071 29 5079.

E-mail address: kurt.kohn@uni-tuebingen.de.

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)² 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).

This approach can be seen as a methodological revision of the original data-driven learning approach advocated to use concordance data from genuine text corpora for the creation of language learning tasks and exploratory activities (Johns 1991; Gavioli 2005). While the concept of data-driven learning originally refers to the pedagogical exploitation of corpora that were intrinsically designed for linguistic description purposes, BACKBONE takes an altogether different route. Incorporating Widdowson's (1990, 2003) imperative of pedagogical mediation, the project emphasises the fundamental need for a thorough pedagogic conceptualisation of the entire corpus creation and deployment process.

Forerunners of the BACKBONE pedagogic corpus approach are the ELISA corpus³ (cf. Braun 2005, 2006, 2007) and the SACODEYL corpora⁴ (cf. Hoffstaedter & Kohn 2009; Pérez-Paredes & Alcaraz-Calero 2009; Widmann, Kohn & Ziai 2008).

2. 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)⁵

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.⁶

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.

² see <http://www.uni-tuebingen.de/backbone> [Accessed 30 Oct 2009].

³ see <http://www.uni-tuebingen.de/elisa/> [Accessed 30 Oct 2009].

⁴ see <http://sacodeyl.inf.um.es/sacodeyl-search2/> [Accessed 30 Oct 2009].

⁵ see <http://www.um.es/sacodeyl/> [Accessed 30 Oct 2009].

⁶ 50 interviews of 10 min. each for English and ELF; 25 interviews of 10 min. each for French, German, Polish, Spanish and Turkish.

3. Pedagogic corpus design, creation, and search

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.

4. From pedagogic corpus search to relevant language learning

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.

5. Concluding remarks

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 authors 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.
- Gavioli, L. (2005). *Exploring corpora for ESP Learning*. Amsterdam: John Benjamins.
- 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* (pp. 291-303). Trier: WVT.
- Johns, T. (1991). Should you be persuaded: Two examples of data-driven learning. In T. Johns & P. King (Eds.), *Classroom Concordancing* (pp. 1-13). Birmingham: University of Birmingham (English Language Research Journal 4).
- 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* (pp. 157-174). Valencia: Universidad Politécnic de Valencia.
- Kohn, K. (2009). Computer assisted foreign language learning. In K. Knapp & B. Seidlhofer (Eds.), *Foreign Language Communication and Learning. Handbooks of Applied Linguistics, vol. 6* (pp. 573-603). Berlin: Mouton de Gruyter.
- Pérez-Paredes, P. & Alcaraz-Calero, J.M. (2009). Developing annotation solutions for online data driven learning. *ReCALL* 21/1, 55-75.
- Widdowson, H.G. (1990). *Aspects of Language Teaching*. Oxford: Oxford University Press.
- Widdowson, H.G. (2003). *Defining Issues in English Language Teaching*. Oxford: Oxford University Press.
- Widmann, J., Kohn, K. & Ziai, R. (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.