

Work package 2.4

System Development & Implementation Integration and Delivery

Deliverable A of WP 2.4

SACODEYL Corpus-Based Language Learning

– Guidelines for Teachers and Learners –

Kurt Kohn (University of Tübingen)

Petra Hoffstaedter (STC)



Minerva



This 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 herein.

SACODEYL Corpus-Based Language Learning – Guidelines for Teachers and Learners – *

What is SACODEYL.....	3
Pedagogic context.....	3
Pedagogic corpora.....	5
The SACODEYL interviews.....	5
Customised annotation	7
Enrichment resources.....	9
Search and display.....	12
Pedagogic integration.....	14
Links to tools and resources.....	16

* **Based on Petra Hoffstaedter and Kurt Kohn (2009).** Real language and relevant learning activities. Insights from the Sacodeyl project. In: Anton Kirchhofer and Jutta Schwarzkopf (eds.) (2009). *The Workings of the Anglosphere. Contributions to the Study of British and US-American Cultures*. Trier: Wissenschaftlicher Verlag Trier.

What is SACODEYL

SACODEYL focuses on the compilation and pedagogical exploitation of spoken interviews with British, French, German, Italian, Lithuanian, Rumanian and Spanish adolescents between 13 and 18 years of age.

The interview transcripts are stored in online corpora and pedagogically annotated and enriched for language learning and teaching purposes. Special emphasis is given to the design and development of pedagogically motivated corpus tools for corpus transcription, annotation, enrichment and search.

To further enhance and explore the corpora's pedagogical quality, corpus-based learning activities are pedagogically embedded in an e-learning environment facilitated through Moodle. Appropriate pedagogic instructions and model courses enable teachers to set up and support corpus-based blended language learning activities combining pedagogically selected interview sections and enrichment resources with communicative interaction among learners and teachers.

The Sacodeyl tools and materials are used by teachers to facilitate pedagogic scenarios guided by the principles of autonomous learning, collaboration and authentication; they are intended for integration into regular classes as part of a more comprehensive and diversified pedagogic approach. These guidelines are therefore in the first place targeted at teachers. They will, however, be also relevant for language learners who are themselves involved in using the Sacodeyl corpus tools for creating their own customised corpora as part of e.g. collaborative learning projects.

Pedagogic context

Modern approaches to foreign language learning and teaching are shaped by the communicative paradigm shift, which has been unfolding and gaining ground in the English language teaching community since the early 1970s. Three insights have been at the centre of pedagogical change:

(a) Learners have been recognised as the real players and decision-makers as to what should go on in class, which has led to a shift from teacher-centredness to learner-centredness and an emphasis on learner autonomy.¹

(b) It has become clear that a more communicative approach also requires more authenticity, in particular in reference to the authentication of learning materials and learning activities in meaningful and relevant discourse events.²

¹ cf. Phil Benson (2001). *Teaching and Researching Autonomy in Language Learning*. London: Longman.

² cf. H.G. Widdowson (2003). *Defining Issues in English Language Teaching*. Oxford: Oxford University Press, 93-109.

(c) Communicative language learning has been identified as an interactive and collaborative undertaking.³

These insights have become constitutive for more advanced versions of communicative language teaching (CLT) like task-based or project-based language learning⁴; later on they were subsumed under the umbrella of constructivism and constructionist approaches to language learning⁵. As a consequence, the objective and challenge of modern language teaching has been seen to lie in facilitating learner autonomy, authentication and collaborative interaction.

Successful facilitation of authenticated language learning greatly depends on the availability of suitable learning materials. In this context, the storage, processing and dissemination potential of web-based digital language corpora can be of great help. Language corpora are, in the first place, large collections of texts compiled according to certain selection criteria and often described and annotated with regard to purpose-driven linguistic categories (“tags”). Lexical and category-based search procedures enable users to quickly locate and extract certain words, word combinations and word contexts (“concordances”) as well as parts of speech or structural units, depending on the kind of annotation.

Up until fairly recently, language corpora were mainly designed and compiled as large representative corpora aiming at the linguistic description and analysis of certain language varieties based on real data. They have enjoyed practical relevance since the publication of the *Collins COBUILD English Language Dictionary* (1987). Throughout the 1990s, corpus-based lexicography flourished and was complemented by corpus-based grammar writing ranging from the *Collins COBUILD English Grammar* (1990) to the *Longman Grammar of Spoken and written English* (1999). With the dramatically increasing storage and processing capacities of personal computers and web servers, corpora and corpus techniques have finally become available to ordinary end-users themselves, including in particular language teachers and learners.⁶

In the following sections, the SACODEYL approach to using corpora for foreign language learning purposes will be described and illustrated. Our approach has a certain “Do it yourself” (DIY) quality: based on results from the European research and development project SACODEYL, we intend to demonstrate how language teachers can be enabled to create and pedagogically exploit their own

³ cf. David Nunan (ed.) (1992). *Collaborative Language Learning and Teaching*, Cambridge University Press.

⁴ cf. N.S. Prabhu (1987). *Second Language Pedagogy*. Oxford: Oxford University Press.

⁵ cf. Dieter Wolff (1994). “Der Konstruktivismus: Ein neues Paradigma für die Fremdsprachendidaktik?“. *Die Neueren Sprachen* 93, 407-429; Bernd Rüschoff (1999). “Construction of Knowledge as the Basis of Foreign Language Learning“. In: Bettina Mißler and Uwe Multhaup (eds.). *The Construction of Knowledge, Learner Autonomy and Related Issues*. Tübingen: Stauffenburg Verlag, 79-88.

⁶ cf. Sabine Braun, Kurt Kohn and Joybrato Mukherjee (eds.) (2006). *Corpus Technology and Language Pedagogy*. Frankfurt/M.: Peter Lang.

small DIY corpora,⁷ focussing on genres and topics of immediate relevance to their specific target group.

Pedagogic corpora

What makes a corpus a pedagogic one? This question points beyond the mere extraction and pedagogic application of concordances, word lists and frequency counts.

Corpora provide access to real language and thus enable teachers and learning content developers to base their pedagogical decisions on empirical facts. However, corpora are also limited in a crucial way: they contain language data which are real but at the same time stripped bare of any contextual information. In Henry Widdowson's understanding, corpora contain texts in the sense of linguistic surface manifestations of discourse.⁸ For successful language learning, it is necessary to involve the learner in such a way that these corpus texts are brought back to some kind of discourse life again.

A corpus is thus not in itself pedagogic; rather, it assumes its pedagogic quality in meaningful and relevant pedagogic discourse. Particularly suitable for this purpose are small corpora which are pedagogically annotated and enriched to meet the thematic and genre-specific language learning needs of a certain target group.⁹

SACODEYL takes up this approach. Pedagogical corpus adaptation includes a pedagogically motivated annotation as well as integration of pedagogically relevant enrichment resources. The overall objective is to help teachers and learners proceed from decontextualised textual data to context-embedded discourse interaction and thus to facilitate and promote authenticated language learning.

The SACODEYL interviews

Each of the small SACODEYL corpora – English, French, German, Italian, Lithuanian, Rumanian and Spanish – contains 20 to 25 video-recorded interviews of about 10 minutes each. Most of the interviews were conducted with individual pupils, some with pairs of pupils.

In order to ensure thematic comparability, a common set of questions was used, covering a wide range of topics including personal information, home and family, present and past living routines,

⁷ cf. Chris Tribble (1997). "Improvising Corpora for ELT: Quick-and-Dirty Ways of Developing Corpora for Language Teaching". In: Barbara J. Lewandowska-Tomaszczyk and Patrick J. Melia (eds.). *PALC-97: Practical Applications in Language Corpora*. Lodz: Lodz University Press, 106-117; Guy Aston, "The Learner as Corpus Designer". In: Bernd Kettemann and Georg Marko (eds.). *Teaching and Learning by Doing Corpus Analysis*. Amsterdam: Rodopi, 9-25.

⁸ cf. H.G. Widdowson (2003). *Defining Issues in English Language Teaching*. Oxford: Oxford University Press, 75-91.

⁹ cf. Sabine Braun (2005). "From Pedagogically Relevant Corpora to Authentic Language Learning Contents". *ReCALL* 17/1, 47-64.

hobbies and interests, holidays, school and education, job experiences, plans for the future as well as open discussion topics. In addition, the interview questions were designed to elicit responses that were likely to favour certain proficiency levels as defined by the Common European Framework (CEF levels) as well as pedagogically relevant language properties.

The following extracts serve to illustrate some of the language learning potential contained in the interviews.

Extract 1 is taken from the English corpus. It concerns giving personal information on CEF level A1/A2 (“I can describe myself, my family and other people”). Pedagogically relevant linguistic properties include the simple present of “be” and “have” and of the general verb “go”, the comparative construction “younger than me”; colloquial usage (“younger than me”, “me and my brother”) as well as useful phrases (“I go to Reading School”, “[we] get on fine”).

Interview extract 1:

Interviewer: Ok, can you tell me something about yourself?

David: Ah I'm fifteen I go to Reading School and I have a brother and a sister.

Interviewer: And are the brother and sister are they older or younger?

David: Ah they're both younger than me. My brother is about a year younger than me and my sister is about three years younger than me.

Interviewer: Do you get on well with them?

David: Yeah overall. I mean I have problems with my sister sometimes but me and my brother get on fine.

Extract 2 is also taken from the English corpus and focuses on past holidays; the CEF level is A2 (“I can describe past activities and personal experiences”). The pedagogically relevant linguistic properties range from simple past (“went”) and present perfect (“I've had”, “I have done”) to prepositions (“went on a rugby tour”), phrasal verbs (“go away”) and idiomatic phrases (“without a shadow of a doubt”).

Interview extract 2:

Interviewer: So would you say your skiing holidays are your best holidays?

Ben: [...] and, you know, I've had I've had family holidays and then I've had holidays with my dad and then I also went away in 2006. I went on a rugby tour with twenty six people from my school. We went to South America. We went to Brazil, Argentina and Chile. And that was without a shadow of a doubt the best thing I've ever done.

Extract 3 is taken from the German corpus. It concerns the discussion topic “all-day school”; the CEF level is B2 (“I can explain a viewpoint on a topical issue giving the advantages and disadvantages of various options”). Pedagogically relevant linguistic properties include subjunctive (“hätte”, “wäre”, “sollte”), hypotactic constructions with conjunctions (“obwohl”, “weil”, “deswegen”) and particles (“eigentlich”, “fast”, “doch”, “gerade”).

Interview extract 3:

Interviewer: Ihr habt ja nun wirklich sehr viel Nachmittagsunterricht. Was würdet ihr dann von der Einführung der Ganztagschule bei uns halten?

Lukas: Also obwohl ich eigentlich fast jeden Nachmittag in der Schule bin, bin ich der Meinung, dass es gerade für die älteren Schüler nicht besonders viele Vorteile hätte, weil wir auch die Möglichkeit haben, in die Stadt essen zu gehen oder uns mit Freunden zu treffen und man doch durch die Ganztagschule 'n bisschen mehr gezwungen wäre, in der Schule zu bleiben und dort zu essen. Deswegen bin ich der Meinung, dass es gerade für die älteren Klassen jedem selbst überlassen bleiben sollte, was er in seiner Mittagspause macht.

The interview format thus proves highly suitable for eliciting spoken language material that combines up-to-date topics with pedagogically relevant linguistic means of expression from grammatical structures to words and phrases, elements of spoken discourse and special pronunciation features. However, pedagogically appropriate corpus tools are needed for making this material accessible for pedagogical exploitation.

Customised annotation

A SACODEYL corpus contains orthographical interview transcripts in XML format, which are structured in short thematic sections and annotated with regard to pedagogically relevant characteristics regarding e.g. topic, grammar, lexis, discourse markers and CEF level.

While a typical corpus annotation is based on words and phrases, annotation in SACODEYL applies to the thematic sections specified during transcription. An interview transcript is loaded into the SACODEYL Annotator, and each interview section is then annotated by selecting categories from a list of available annotation categories and assigning them in a drag-and-drop fashion to the respective section. The specific parts in a section to which a category applies (i.e. a word form or grammatical construction) are marked by a colour coding.

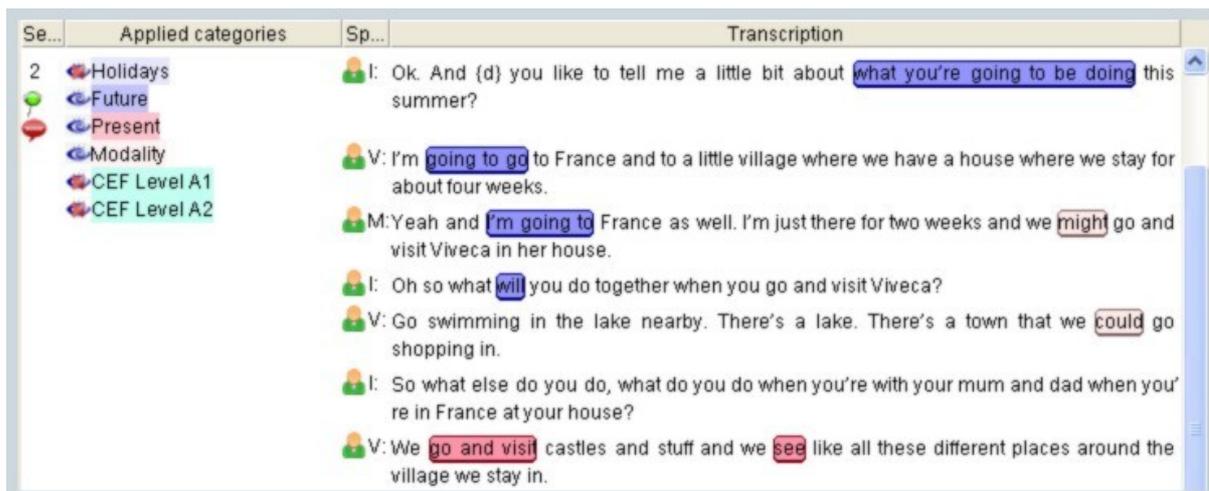


Fig. 1: The SACODEYL Annotator

In this way, transcripts can be annotated in an efficient manner without any specialised computer knowledge or skills. And what is more, the SACODEYL Annotator also enables teachers to define their own customised annotation categories (see Fig. 2).

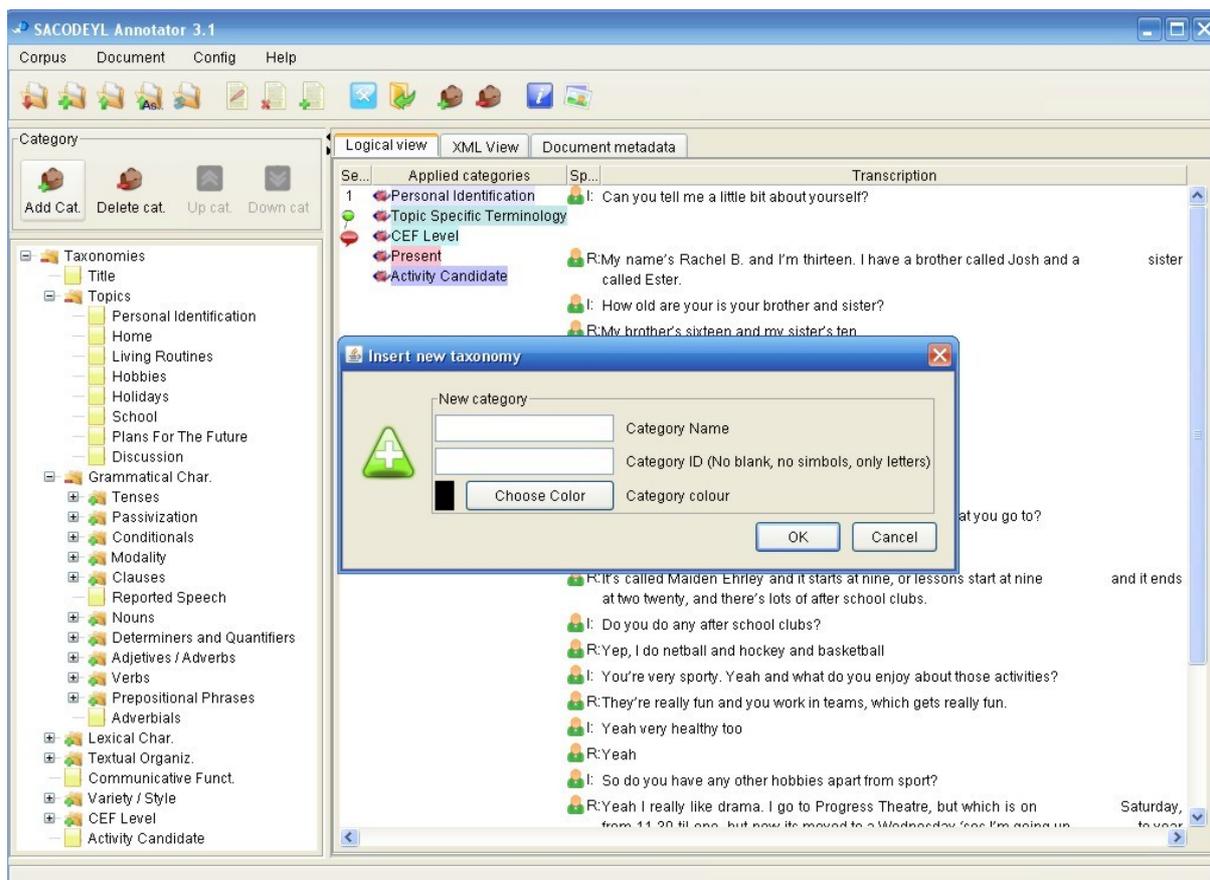


Fig. 2: The SACODEYL Annotator – Customisation of annotation categories

This allows them to annotate the interview transcripts with regard to those characteristics they deem pedagogically most relevant for a particular target group or learning purpose.

Enrichment resources

The SACODEYL corpus data represent pedagogically annotated but otherwise decontextualised interview transcripts. Further pedagogic mediation materials and procedures are thus required to help teachers and learners authenticate these data in meaningful and relevant pedagogic discourse activities. For this purpose, access to various web-based enrichment resources is provided, including the interview sound and video files, cultural information, ready-made web-based language learning tasks and instructions for explorative and communicative learning activities.

Enrichment resources are stored in the Virtual Resource Pool (VRP) in the form of web links consisting of a name, type specification, brief description and web address. In the process of annotating a corpus, relevant resources were selected from the VRP and pasted into Resource Sheets (RS); these were then attached to the respective interview section for enrichment. When searching a corpus, users can access available resources via these resource sheets.



Fig. 3: Virtual resource sheet with resources for an interview section

The **language learning modules** forming part of the SACODEYL enrichment resources are created with Telos Language Partner¹⁰, a template-based authoring tool that supports a wide range of language learning specific activities. Depending on their inherent language learning potential, interview sections are used in language-focused Telos tasks, which may combine lexico-grammatical and pragmatic knowledge development with listening, reading and writing exercises. The basic materials in these tasks include excerpts from interview transcripts as well as sound or video files or any other enrichment resources that meet the respective learning objectives.

¹⁰ cf. Kurt Kohn, "Telos Language Partner: Do-it-Yourself Authoring for Content-Based Language Learning". In: Ana_M. Gimeno (ed.). *Computer Assisted Language Learning: Authoring Tools for Web-Based CALL*. Valencia, Universidad Politécnic de Valencia, 157-174; also see the Telos learning packages on <http://www.sprachlernmedien.de> (31 May 2008) [→ Open Learning Space].



Fig. 4: Exercises from a Telos learning module

The Telos tasks are available for corpus users as ready-made learning units; but they also serve as models for teachers who want to produce their own customised exercises.

To get a free licence of the Telos Language Partner software, go to “Authoring Software” at

www.sprachlernmedien.de

The **explorative and communicative exercises** are specified in PDF sheets. They put special emphasis on integrating interview-related learning activities in a classroom and/or Moodle-based e-learning environment. A typical exercise might combine a corpus search with a forum discussion, a wiki collaboration or a group interaction in class.

Specific interview topics might be taken as a starting point for further exploration and discussion activities that invite pupils to expand their cultural and linguistic knowledge individually and in cooperation with others and to practise their reading and speaking skills as they go along. In connection with a relevant interview (section), these activities provide opportunities for authenticated learning experiences.

The following examples are taken from the English corpus enrichment resources.

TOPIC 1: Who I am and where I live	
Exploratory Exercise	
CEF level: A2	
Task topic 1 – E10	
Activity	The students explore the extracts of the corpus where the interviewees talk about a number of English cities. They look for these cities on a map and find additional information about them.
Task	<ul style="list-style-type: none"> Go to the “co-occurrence search” tab. Search for one of the following places: Aurora, Bristol, Oxford, London, Jersey, Liverpool, Cardiff, Sheffield, Bracknell, Bournemouth, Cambridge. Find out why the interviewees mention these places and summarize it on a piece of paper.

	<ul style="list-style-type: none"> ● Go to http://wikipedia.org and see what else these places are famous for. ● Add some information that you find interesting to your summary. ● Add a sketchy map to show where on the British Isles (except for Aurora) this place is. ● Put your paper on the wall and read what the others have found out.
Presentation of the results	The pupils must read out their summary to the class.
Hint	If the group is large, you might work in small groups instead of individually so that each of the places is assigned to one group.

TOPIC 1: Who I am and where I live	
Communicative Exercise	
CEF level: A2	
Task topic 1 – C5	
Activity	<p>Talking about moving your home.</p> <p>The students explore the extracts of the corpus where the interviewees talk about what happened when they moved. They relate it to their own experience and describe what happened in their own or in their friends' lives when they moved.</p>
Task	<ul style="list-style-type: none"> ● Go to the “co-occurrence search” tab. ● Search for “moved” in combination with the sections “Personal Identification” and “Home”. ● Set “search scope” to “section”. ● Read through the results and watch those sections that you find interesting. ● Take notes about what happened to your own life when you moved to a different town or – if you have never moved – what happened to a friend of yours when he or she moved to a different town.
Presentation of the results	<p>Each pupil has to present their story of “moving” to the class.</p> <p>If appropriate, the students should add their story to their profile on Moodle or put it into some other place where the pupils present themselves (e.g. the classroom wall).</p>
Hint	<p>This exercise is a suitable follow-up activity for task topic 1-E9.</p> <p>If you have a class with very different cultural backgrounds, you might limit the story to a certain aspect of moving, e.g. about the differences in the new school.</p>

TOPIC 2: Leisure Activities	
Explorative exercise	
CEF level: B1, B2	
Task Topic 2-E6: Portraying your favourite bands	
Activity	The pupils have to find information on English web sites about current English bands and they compile a portrait of their favourite band.
How to go about	<ul style="list-style-type: none"> ● First, the class decides jointly which features of the group all pupils have to search for (e.g. age, musical inspiration, how the band got together, etc.). ● Each pupil sets out to find the information on the web regarding his/her chosen singer/band. ● The pupils write down the information on a piece of paper. ● Homework: Each pupil has to take the information and put it online in a wiki, together

	with the song text of one of their favourite songs.
Presentation of the results	<ul style="list-style-type: none"> • Wiki in Moodle
Web sites	http://en.wikipedia.org/wiki/Category:English_musical_groups

All enrichment resources (Telos learning modules and the explorative and communicative exercises) can be accessed via the SACODEYL Search Tool:

<http://purl.org/sacodeyl/searchtool>

To access the resources, select a corpus and then click on “Resources” in the menu.

Search and display

A corpus and the attached enrichment resources can only be as good as the procedures that are available for finding and processing the information and materials it contains. The SACODEYL search & display interface operates online and supports a flexible combination of pedagogically motivated options.

The BROWSE function enables users to scan the entire interview corpus, providing (if available) a picture and a short description for each interview as well as access to the interview transcripts and video clips. This offers teachers and learners the opportunity to read the corpus in a linear fashion, like an ordinary collection of texts, with the additional advantage of a discourse experience. BROWSE is particularly suitable for small corpora.



Fig. 5: The SACODEYL search tool – BROWSE interviews

SECTION SEARCH is based on the categories used during the annotation process. Users select any combination of thematic and linguistic annotation categories from a list and then choose between searching for sections that fulfil at least one of these categories (“or” condition) or all of them (“and” condition). It is also possible to restrict a search to the set of sections that has resulted from a previous search.

Pedagogic integration

In SACODEYL, the actual pedagogic exploitation of the annotated and enriched interview corpora is further enhanced by a Moodle-based e-learning environment. Its main function is to help teachers provide and monitor authenticated learning opportunities that combine language-focused Telos tasks with collaborative learning and communicative interaction.

It is important to note that learning is conceptualized and organized as a pedagogic blend of online and offline, classroom and homework activities. The e-learning environment assumes a double function in this connection: it facilitates online learning and at the same time is used to orchestrate the entire pedagogic event.

To create customized blended learning units for their target groups, teachers can choose activities from the various SACODEYL corpus resources and combine them as pedagogically required:

First, the available Telos learning modules support fairly closed ready-made learning activities based on selected interview sections. They cover the whole range of topics addressed in a corpus and focus on linguistic forms, functions and skills. The Telos exercises are intended to offer pedagogically structured learning opportunities that help students to work with the interviews in particular with regard to the challenge of understanding textual manifestations of spoken discourse and practicing relevant linguistic means of expression.

Second, the suggestions for explorative and communicative activities adopt a more open approach. Communicative tasks often start with listening to interview sections covering a specific topic, and are followed up by group or partner interaction in class or collaborative e-learning activities using the forum or wiki functions in Moodle. Explorative tasks include web searches as well as activities where students are asked to use the SACODEYL search tool to find out what different speakers say about a certain topic or which linguistic means of expression they are using. The tasks have a rather flexible design and can easily be adapted by teachers to the specific needs of their students.

The following example illustrates how teachers can combine the available materials and use them in their language courses. The example is based on the French corpus and addresses learning objectives characteristic of CEF level A2.

Se présenter



Lernmodul:

[Mélissandre et Thibault habitent à Gosier en Guadeloupe](#)

Was weißt du über Guadeloupe?

Geht auf die unten angegebene Webseite und versucht, die folgenden Fragen zu beantworten.

- Wo liegt Guadeloupe?
- Wie weit ist die Guadeloupe von Frankreich entfernt?
- Was ist die Amtssprache auf Guadeloupe?

[Guadeloupe - Wikipedia](#)

Cherchez des extraits de vidéos dans lesquelles les élèves se présentent:

Vorgehen:

1. Auf den Link klicken, um die [SACODEYL-Suchsoftware](#) zu öffnen.
2. French Corpus auswählen.
3. Auf "Search" klicken.
4. "Section Search" auswählen.
5. Die Kategorie "Présentation personnelle" (Unterkategorie zu "Sujets") auswählen.
6. Schaut euch die angezeigten Interviewausschnitte an und sucht Ausdrucksmittel heraus, mit denen man sich vorstellen und etwas über sich erzählen kann.

Und nun könnt ihr euch selbst vorstellen und etwas über euch erzählen.

Erwähnt dabei eine Sache, die nicht zutrifft. Die übrigen Schüler sollen herausfinden, welche Information falsch ist.

[Forum - Se présenter](#)

Fig. 7: Pedagogic integration in Moodle

The unit starts off with a Telos learning module based on interview sections in which two 13-year-olds, a boy and a girl, from Guadeloupe introduce themselves and talk about things they like. The learning module offers a structured self-study unit introducing the topic and offering listening practice and comprehension exercises. In addition, learners are familiarised with the basic means of expression enabling them to introduce and talk about themselves according to A2 level requirements. The module also includes grammatical exercises for practicing the conjugation of verbs in the present tense.

All the activities that follow build on the Telos learning module offering various opportunities for more or less authentic exploration and communication activities. Since the two interviewees introducing themselves in the Telos module are both from Guadeloupe, learners are given a simple explorative task requiring them to search a website for basic information on Guadeloupe; an

appropriate web link is provided along with some guiding questions. In another explorative task, the learners are asked to use the SACODEYL corpus search to find out how different speakers introduce themselves and to analyse the linguistic means of expression that are used in relevant interview sections.

Performing the explorative tasks in pairs of two or in small groups adds a collaborative and communicative dimension. Learners could, for instance, be asked to note down information on the speakers they focus on, compare and discuss thematic and linguistic peculiarities in a group forum and give an oral presentation later in class. By inviting learners to introduce themselves in a multimedia blog, an even deeper personal involvement and authentication would be achieved. These activities would find their natural continuation in e-twinning projects between schools from different countries.¹¹

A SACODEYL Moodle Demo Course is available at

www.sprachlernmedien.de

Pedagogical evaluation in schools is needed to shed light on the potential and feasibility of the SACODEYL approach. This concerns in particular the nature of the learning and teaching activities involved, teacher and learner satisfaction as well as learning outcomes. Due attention should also be given to the media competences required by teachers and pupils and to the need for continuous training. Additional attention should be given to the overall definition and organisation of the teachers' workload. While e-learning promises to open up new spaces for autonomous collaboration and authentic interaction, these promises will only come true if teachers are enabled to provide pedagogic guidance and support outside and beyond the traditional classroom environment.

Links to tools and resources

SACODEYL Annotator	http://www.um.es/sacodeyl
SACODEYL Search Tool:	http://purl.org/sacodeyl/searchtool
Telos Language Partner software	www.sprachlernmedien.de
Telos learning modules and the explorative and communicative exercises can be accessed via the SACODEYL Search Tool. To access the resources, select a corpus and then click on the Resources menu.	http://purl.org/sacodeyl/searchtool
SACODEYL Moodle demo course	www.sprachlernmedien.de

¹¹ ><http://www.etwinning.net>> (31 May 2008).