

# NLP serving the cause of language learning

**Frédérique Segond**

Xerox Research Centre Europe  
6, chemin de Mauperthuis  
F-38240 Meylan - FRANCE  
segond@xrce.xerox.com

**Thibault Parmentier**

Xerox Research Centre Europe  
6, chemin de Mauperthuis  
F-38240 Meylan - FRANCE  
parmentier@xrce.xerox.com

## Abstract

E-learning paves the way to a new type of course, more student centred, granulized, on demand, and highly interactive. Natural Language Processing (NLP) technologies associated with other multimedia technologies can help to address the major issues raised by this new type of courses: interaction, personalization and reliable information access. This paper presents Exills, a true e-learning solution which integrates natural language processing tools and virtual reality<sup>1</sup>. Exills is unique in that, unlike most of the language learning systems, it focuses on improving learners' performance rather than learners' competence.

## Introduction

This paper is not a theoretical paper. It describes a true e-learning system that concentrates on improving users' performance in a foreign language and integrates natural language processing technologies.

The system, Exills ([www.exills.com](http://www.exills.com)) represents an innovative way of integrating Natural Language technologies and multimedia technologies in order to provide a visionary e-learning tool for learning foreign languages.

Exills is the result of a close collaboration between language teachers, specialists in virtual reality, linguists and computer scientists. While retaining "traditional content" such as language exercises, grammar worksheets and speech acts, Exills takes advantage of new technologies for interaction (asynchronous and synchronous), game aspect and course scenarization.

Exills has been conceived with the constant concern for pedagogical quality and for differentiating itself from both CDROMs and face to face language courses. We have been careful to

renounce using technologies, even innovative ones, if they did not serve the cause of a pedagogical e-learning system. The focus has been to improve learner's performance rather than to insist on improving learner's competence<sup>2</sup>. Indeed, experimenting with different available language learning solutions focusing on competence (via vocabulary exercises and grammar) led us to the conclusion that these types of activities do not insure that learners are then able to interact in a foreign language. While they might acquire the conceptual knowledge of the language they are not necessarily able to use it in practical contexts. In the case of Exills, learners are put into practical situations. The situations are represented via different scenarios. In most cases the scenario corresponds to everyday work situations.

Exills focuses on the interaction aspect of language learning driven by the use of the Internet, in particular, on interaction and information access.

Interaction within Exills includes language understanding through reading and listening. It also improves language production in writing. As we will see in the conclusion adding speech technologies to Exills will enable users to also improve on language oral production.

In the demonstration scenarios that are accessible online, targeted users are professionals with an intermediate level in English or French. These scenarios put users in typical work situations such as introducing themselves, reading emails, searching for and understanding information, writing and making a presentation, ordering their meal, interacting with their colleagues and customers. Part of the scenario is dedicated to the discovery of cultural aspects.

---

<sup>2</sup> Competence and performance as defined by Chomsky (Chomsky, 1965): *Competence* is the speaker-hearer's knowledge of his language. *Performance* is the speaker-hearer's actual use of his language in concrete situations

---

<sup>1</sup> The software we use for Virtual reality comes from Blaxxun (<http://www.blaxxun.com/en/site/index.html>).

In what follows, after presenting our view of e-learning, we describe Exills and its different technical components. We then conclude by suggesting other linguistic technologies that could serve the cause of language e-learning.

### **How e-learning affects our way of learning**

Learning through the Internet, or e-learning, paves the way for a new type of course which, most of the time, complements the face-to-face course (blended learning), but can also be substituted for the face-to-face courses. While traditional and Computer-Based training is targeted to a group of students, e-learning is often referred to as “my-learning” as it is highly personalized, granularized, on demand and highly interactive. The characteristics of e-learning are that it provides learners with the possibility to learn exactly what they need, from anywhere and at any time

E-learning completely changes both the teaching and the learning processes. Teachers become tutors that support students’ work. Students take more responsibility in the learning process as they know what they need to learn to be efficient. This is especially true in a life long training context. They become actors in their learning process by choosing their course topics, elaborating their learning paths, exchanging with other students as well as with tutors.

E-learning is definitively a new way of learning that can take advantage of new Internet technologies.

However, despite the explosion of e-learning platforms also called LMS for learning management system (see [thot.cursus.edu](http://thot.cursus.edu)) there are very few systems which take real advantage of the functionalities the Internet offers such as: interaction, personalization and huge information access (for both students and tutors). Indeed, most of the current e-learning solutions are LMS that are former Knowledge Management Systems customized for learning. Because they have not firstly been designed for a learning purpose they often miss the pedagogical dimension. Other solutions are a simple web link to already existing CD-ROM content. This solution usually offers a simple type of interaction such as emails with the tutors and, sometimes, forums with other students.

Within the e-learning domain, language learning is one of the most successful market segments as it comes second only to IT and marketing. This is probably due to the fact that for many years it has been possible to learn a language using electronic media (CD-ROM). Medium to large publishers have been successful in developing language training multimedia content

of good quality. Still, the majority of so-called e-language learning solutions that are on the market today are CD-ROM content transferred onto the web rather than real online solutions.

The reason is probably that most of the publishers in this area are specialists in producing good quality pedagogical content but are not specialists of the Internet, nor do they have the technological resources necessary to transfer their content onto the Web. Offering on-line content on the web is very different from offering content on a CD-ROM. These media organize content in a completely different way and this has a strong impact on the basic pedagogical principles. This shows up at different levels. For instance, studies have shown that users will spend at the most 15 minutes on an Internet learning session versus 45 minutes on a CD-ROM. learning. Besides, there is the issue of how to organize learning material so that navigation among short learning modules is possible without affecting the learning pedagogical process. Finally, record keeping, tracking, tracing and flexibility are all essential functions for making an on line course successful as well as for keeping user motivation high. These are just a few of the issues that need to be addressed when creating high quality content on the Internet for language learning.

Exills is a true e-learning solution, that on the technological part, integrates virtual reality with linguistic and collaborative technologies as well as with a smart search engine.

### **Exills’ general features**

In very general terms, Exills could be described as a hybrid between an authoring tool and a delivery platform particularly well suited to building and distributing language learning applications.

Exills user interface provides interactive access to the most up-to-date Internet technologies.

The user interface is composed of:

- a 3D scene to put the user in a cognitive context in line with the topic of the course;
- a chat frame to allow students, robots and tutors to interact via text (written production);
- a frame to provide learners with any information they might need as well as with activities to do.
- a toolbar to provide learners with easy access to specific functionalities such as: a linguistic toolbox (contextual dictionary look up, phonetic spell checker, conjugation

engine), a notepad to take notes, a history of their work to review it at any time, as well as shortcuts, through maps, to navigate inside 3D scenes. They can also get help from tutors, access their profile and a list of personalized exercises, grammar worksheets, and speech acts.

Each course is associated with a scenario. These scenarios are usually associated with a mission to accomplish. This mission integrates learning activities, game aspects and competition between students. In order to achieve the mission, learners need to interact with each other, with the tutor as well as with the virtual environment. Thanks to the virtual reality this interaction takes place like in video games. Learners play a video game in a foreign language.

In its current incarnation Exills has the following underlying features.

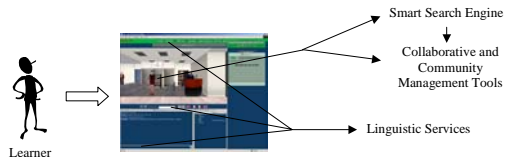


Figure 1: Technologies involved

### Virtual reality

While virtual reality presents real advantages for learning, this technology is not yet strongly used for this purpose.

In Exills, virtual reality is used both for interaction and for personalization. Personalization, because when they are in virtual reality, learners can choose a representation of themselves, also known as *an avatar*, and can move around the virtual places like in the real world. For instance, learners can meet with other avatars/learners, interact with them and access the Internet.

The virtual reality induces a type of interaction that completely differs from email, forum or face to face. Part of this difference is due to the fact that people hide themselves behind an avatar but also because of the game aspect permitted by the virtual reality.

First experiments showed that in virtual reality, as in games, people behave completely differently and tend to interact much more. For instance, we have noticed that while in a face to face or in a video course, most of the time, only the most confident students speak up; in a virtual environment, everyone dares more to express himself/herself. This definitively helps the tutor in evaluating students' ability.

To ease and push learners' interaction, it is important to provide them with an interaction

topic. This is where the notion of scenario comes in and where, again, virtual reality turns out to be an important technology. The scenario is there to help students to interact among themselves but also to immerse them directly into "prototypical situations". For instance, if students need to learn a language for work reasons, a scenario will be built around their everyday life at work, including the physical work place and this induces reaction and discussion correlated to the 3D scenes. Because it provides learners with a reason to interact, like in role playing games, the scenario makes interaction more effective than in traditional chat and forums.

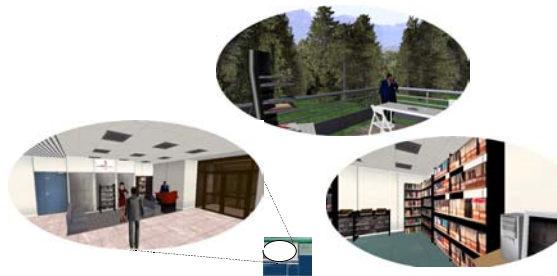


Figure 2: Set the student in a cognitive context using the 3D.

What is more natural than clicking on a computer screen to access emails, than clicking on a CD to hear music, than visiting a library to read books or than being invited to chat with the receptionist when arriving at the reception desk? Because it mimics the real world, students can enter quickly into the virtual reality environment. Students usually just need a few hours' training to easily manipulate the mouse in order to move from one virtual place to another.

### Linguistic technologies

Exills adds other technologies to enhanced virtual reality in such a way that the resulting learning system makes real use of the potential of the Internet.

In its current incarnation, Exills offers learners contextual bilingual dictionaries, morphological analyzers and syntactic taggers, a language identifier and phonetic spell checkers. These technologies are there to help students express themselves or understand a foreign language. Linguistic technologies associated with virtual reality retain the personalized aspect of individual courses while being immersed in the collective course.

Within the chat, Exills offers a comprehension help service which uses the context to first retrieve the most appropriate translation of a word (syntactic and semantic disambiguation). As a

result, the students' comprehension is sped up and interaction becomes smoother.

The same contextual dictionary look up could be used for providing the student with a list of sentences where the word they don't understand appears with the same meaning. In that case, the pedagogical process is different: instead of providing the student directly with the translation, the system provides him/her with other contextual uses of the same word

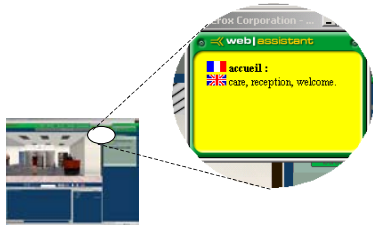


Figure 3: Understanding assistance.

If we now consider the linguistic analysis of student production (chat/forum contribution, exercise, writing, etc.), linguistic technologies provide students with customizable services which allow them to parse and tag their own production in order to check its correctness. Here again the goal is not to provide them with the correct answer, this is left to the tutor, but rather to help them detect their own mistakes.

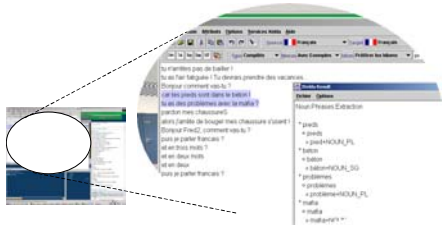


Figure 4: Through linguistic services, students can analyze and correct their written production.

Along the same lines as “learning by example”, linguistic tools can be used to highlight specific language constructions such as functional relations, terminology or idiomatic expressions.

Derivational morphology can be used to discover the words of a same family and improve students' vocabulary.

A simple morphological analyzer gives students access to verb conjugations or noun declensions.

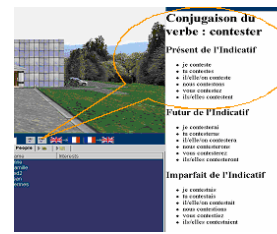


Figure 5: Access to the conjugation of a verb

Exills also includes a language identifier to force students to interact only in the language they are learning. This integration relieves the tutor of checking on each student, and pushes students to express themselves more in that language.

The language identifier is integrated into the chat. This module identifies the language used by a student and if different from the one the student is supposed to learn, the system blocks any type of interaction. Students then receive a message saying that they are not using the appropriate language.

The language identifier can also be plugged into mail or forum services between students or even between students and tutor.

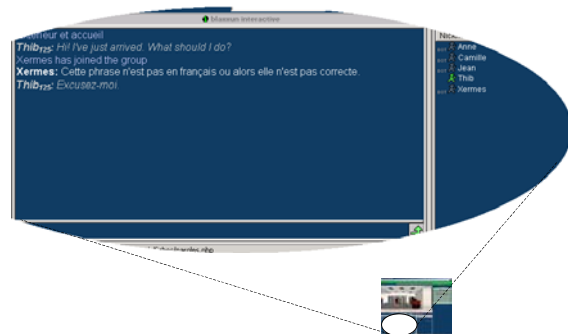


Figure 6: Automatic control of the language used in the chat.

A phonetic spell checker is also available to students to help them to learn the correct spelling of words.

Linguistic technologies can be used to reinforce the personalization aspect of e-learning. Indeed, linguistic technologies open the way to learning on any text that is of interest for the course and for each individual learner. For instance, professionals can work on their everyday technical documentation and can learn their professional terminology in a foreign language. In particular, linguistic technologies can automatically extract technical terminology for them.

### Smart search engines

Exills uses smart search engines in order to strengthen the personalization and interactive aspects. Indeed, having a smart search engine that can take into account students' needs and interest

through a detailed user profile and can retrieve personalized content from the web clearly improves the personalization aspect and course efficiency. Tutors and students can use the Internet as an infinite source of information to enrich course content. This is where collaborative technologies play a crucial role. Information on the Internet varies in quality, and it is especially important in a course context to be able to separate the wheat from the chaff.

Smart search engines strengthen the personalization aspect of e-learning. They can retrieve documents that match the user's profile. This user profile encompasses student-specific difficulties in the courses as well as the student's personal interest, be it for work reasons or in life in general. An example of this is a student who is interested in American politics and has difficulties with the use of idiomatic expressions in French. A smart search engine uses morpho-syntactic analysis of the retrieved text as well as an ontology to provide the student with texts about American politics that contain idioms. Using smart meta-search engines permits a search on the Intranet as well as on student emails. This is especially useful in a context of professional language learning courses. In that case, the search can be performed on technical texts from the company.

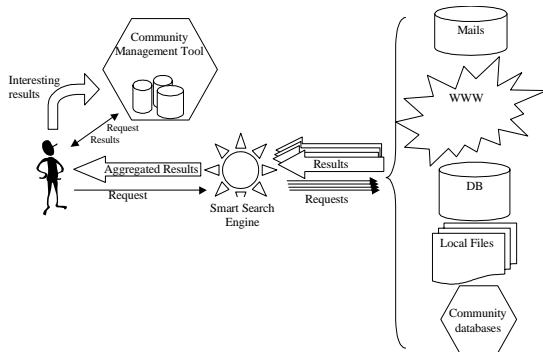


Figure 7: Correlated use of a content aggregation engine and of a community manager tools.

### Collaborative technologies

Within Exills, collaborative technologies are used to filter information found on the web by commenting and interacting on it. At the end of the interaction process, the tutor will decide which documents will be retained to enrich the content of the course.

Using collaborative technologies impacts the two kinds of users in several ways.

- First, the students. Using this tool and contributing to its content allows the student

community to have access to a very qualified and annotated text database.

- Second, the tutors. Tutors can use the community repository in two different ways: first, to get specific and adequate content in order to generate new exercises or activities; second, to evaluate the level and the understanding of the students by analyzing their comments on the different documents posted.

### Infinite number of possible scenarios

The Exills scenario that is currently available on the Internet relates to professionals working for a printing company. It is just one of many possible scenarios. In fact, Exills permits the construction of an infinite number of scenarios. This possibility also points out the high pedagogical adaptability of Exills which leaves to the tutor the possibility to decide about scenario, content and activities. In the next sections we give some examples of other possible scenarios that could use Exills as an underlying framework.

### Schools

In an academic context, it is possible to organize virtual travel for students, similar to language exchanges that already exist. In that, case pupils from two different countries connect at the same time and organize a virtual tour of their school, their district, their city, they describe their cultural habits. In this type of experiment one class plays the role of the tutor and corrects the language of the other class.

This kind of experimentation is known as 'Tandem' in the pedagogical domain. According to several studies, this way of learning languages is more efficient as pupils are more motivated.

### Life long training

One important pedagogical consequence of the notion of scenario is that it allows users to learn both a work process and a foreign language.

Another scenario, we are currently working on, is targeted to people working at hotels reception desk. This scenario is developed in the framework of an eContent European project, Thetis (<http://www.thetis-project.org/>), which groups together a language publisher (Q group<sup>3</sup>), a training company specialized in tourism (Grupo

<sup>3</sup> <http://www.qgroupplc.com/>

GDT<sup>4</sup>) and a research center (Xerox Research Centre Europe<sup>5</sup>).

To summarize, while using linguistic technologies, students can process online texts on the fly and can work on their own documents. This strengthens the personalization aspect of the solution. Going forward, using Exills in a professional context allows students to work on the documents of their own company.

## Evaluation

Exills has been tested in different contexts. We tested the current scenario (professional working for a printing company) in language schools as well as in companies. In all cases we have spent two hours beforehand with the tutors to train them with the concept as well as with virtual reality. Between the training of the tutor and the real test we allowed one week so that tutors can get familiar with the environment. We attended the testing just to help in case of technical problems. We distributed a questionnaire at the end so that users can immediately provide us with feedback. The feedback was globally highly positive. Users like the game aspect, although sometime they would prefer a non professional scenario. Because of this game dimension they tend to use the linguistic technologies more than usual and in particular the contextual bilingual dictionary look-up. This demonstrates that the wish to interact combined with the impossibility of speaking any other languages, because of the language identifier, push them to use linguistic technologies to progress.

Teachers like the language identifier as it relieves them of checking on the language students use. They report on the fact that correcting on the fly is sometimes difficult because the chat is so quick between participants. In any case they do have the possibility of correcting any student production afterwards as everything is stored on the server.

The other context where we tested Exills is within classrooms. In that case users were teenagers from age 13 to 14. This type of population is obviously very comfortable with virtual reality and the game aspect. Again they enjoy very much the fact that they learn a language while playing a game, and again they made extensive use of the linguistic technologies to be able to write sentences. School teachers were impressed by the fact that their students stayed attentive for the entire course. This should

probably be balanced by the novelty aspect of the system.

One thing we have not yet been able to test, but which would be very interesting to evaluate, is the difference in users performance in language after having used Exills.

We are now working with a research team of specialists in pedagogical issue to define a methodology that could measure this performance.

## Conclusion

To summarize, Exills integrates virtual reality, linguistic, collaborative and smart search technologies into a true e-learning solution that can be used both synchronously and asynchronously. Thanks to this technology integration Exills allows:

- Students to benefit from the internet's potential for interaction, to have a controlled and efficient access to the information on the Internet, to learn by themselves at their own rhythm.
- Tutors to get support in supervising their students.
- Learning companies to offer courses based on a new concept (virtual reality and linguistic technologies) instead of just putting classical lessons on the web.
- End-user companies to have a course completely customized to their needs, their working environment and their terminology.

Exills is a true e-learning solution created to take advantage of the web. By taking advantage of the interaction, Exills escapes the drawbacks of many e-learning solutions such as: the strenuousness of reading notices, the impression for students of being left to their own devices, or the temptation to give up the course. This is made possible because:

- Virtual reality induces a cognitive context for the students and pushes them to interact.
- Smart search technologies allow one get information, texts and documents from the Internet or specialized databases related to a specific subject matching the language level of the student.
- Collaborative technologies allow the sharing of information gathered on the Internet between students and the provision of adequate content to the tutor to create new exercises or lessons almost on the fly.
- Linguistic tools provide autonomy to students by showing them concepts, give

---

<sup>4</sup> <http://www.grupogdt.com/superior.html>

<sup>5</sup> <http://www.xrce.xerox.com/>

them assistance to understand word meanings or different senses within a particular context by presenting various examples, to provide feedback for their production (in the chat as well as during exercises or even free activities).

Clearly Exills would benefit from adding new technologies and in particular speech technologies. With speech technologies, students would be able to interact via voice, to record themselves and to compare their own oral production to a model.

Finally, the linguistic technologies presented in this paper could also be used to provide more functionality. For instance, morphology and tagging could be used in order to automatically create online filling-in-the-blank type of exercise from any text found on the Internet: The process is quite similar to the exercises generator framework but the functionalities provided to the student are fewer: given a text and an exercise pattern, the exercise is generated just for her/him. This exercise is then corrected automatically using pattern matching, morphology or spellchecker.

Part of these ideas, especially the speech components<sup>6</sup>, will now be implemented in the framework of Thetis, the on-going European project mentioned before.

## References

- Chomsky N., . *Aspects of the Theory of Syntax*. Cambridge: The MIT Press, 1965.
- Baudit, Apprendre à deux. Études psychosociales de situations dyadiques, PUF, 1998. Paris.
- Bayer V., Farah J., *Apprentissage des langues en Tandem sur Internet*, ELA, n°113, pp.73-78, 1999.
- Bergholz A., Experiences with the knowledge Pump, *ISIC 2002: Information Needs, Seeking & Use*, September 2002, Lisbon.
- Brun C., Parmentier T., Sandor A., Segond F., Les outils de TAL au service de la e-formation in *Multilinguisme et le traitement de l'information*, Frédérique Segond Ed., Hermès 2002, pp. 223-251
- Chidlovskii B., Glance N., Grasso A., Collaborative Re-ranking of Search Results, *Proc. AAAI-2000 Workshop on AI for Web Search*, July 2000, Austin.
- Kashny M., *Les usages des Technologies d'Information et de Communication par des enseignants dans un dispositif de formation tutorée en langues vivantes étrangères. Une approche ergonomique*. Thèse Université Pierre Mendès France, 2001, Grenoble.

Kindley, R., *Scenario-Based E-Learning: A Step Beyond Traditional E-Learning*, <http://www.learningcircuits.com/2002/may2002/kindley.html>

Leacock C., Chodorow M., Miller G., "Using corpus statistics and WordNet relations for sense identification, *Computational Linguistics*, 24(1):147-166, Mars 1998.

Paulsen, J. (2001-section 4). "Authentic Online Target Language Reference Resources." New Era Trends and Technologies in Foreign Language Learning: An Annotated Bibliography - *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, April 2001, Wake Forest University.

Puren, C. *Histoire des méthodologies de l'enseignement des langues*, Clé International, Nathan, 1988, Paris.

---

<sup>6</sup> This will be done using the Wimba technology  
<http://www.wimba.com/index.php>