

The ALLES Project (Advanced Long-distance Language Education System)

Enrique Torrejón

Universidad Europea CEES
Villaviciosa de Odón, 28670 Madrid
enrique.torrejón@uem.es

Celia Rico

Universidad Europea CEES
Villaviciosa de Odón, 28670 Madrid
celia.rico@ti.fil.uem.es

Abstract: ALLES is a new project on Computer-Assisted Language Learning (CALL) for advanced learners of English, German, Spanish and Catalan in the domain of business and economy. NLP tools and techniques will play an important role in ALLES for evaluating user's oral and written production as well as for gathering and tailoring teaching materials.

Keywords: CALL, NLP Tools, Speech Recognition, Parsing Techniques

1 Project Summary

The ALLES Project started in June 2002 and it is funded by the Information Society Technologies Programme from the European Commission (IST-2001-34246). The main objective of the project is the design and development of an advanced long-distance second-language learning system for advanced learners of English, German, Spanish and Catalan in the specific domain of business and economy.

The project will be carried out by the following participants: SchlumbergerSema as project coordinator, Universidad Pompeu Fabra, Institute of Applied Information Sciences from Saarbrücken, Germany, University of Utrecht from The Netherlands, and Universidad Europea CEES.

The resulting system will include learning modules for listening comprehension, reading comprehension, written and oral production as well as a supporting tool containing a grammar and a dictionary. Special emphasis will be placed on developing adequate NLP techniques, particularly in the written and oral production modules, that could render relevant feedback to the user according to a carefully thought didactic design which takes into account the special characteristics of e-learning systems.

2 Technical Objectives

The ALLES Project sets out to achieve four different objectives in the field of Computer-Assisted Language Learning (CALL):

2.1 The Effective Didactic Design and Content Selection

In order to avoid inconsistencies in the design of the four different modules corresponding to the four language skills, the didactic design in ALLES will play an important role by specifying the learning sequence in the system, the interaction among the four main modules, the different gradation and modularity of the learning units defined according to specific tasks, and the appropriate feedback information that can ensure the user a good learning path.

Content materials for the system will be selected from different business and economic sources that range from newspapers, online business information, economic reports to pieces from radio and television broadcasts, seminars, specialized talks, etc. These materials will have to be tailored for the different learning modules and will also provide examples for the grammar and the

dictionary that will be embedded in the supporting tool.

2.2 The Practical Use of NLP Tools

Restricting the subject matter to the field of business and economy allows for a controlled context where the use of NLP tools could be feasible and provide the required feedback for correct evaluation of the user performance in an advanced long-distance language learning system. The use of NLP tools in CALL systems has been gaining momentum over the last few years and it should be expected that they will play a central role in the years to come.

In ALLES, the following NLP tools will be very essential for the evaluation of the user production in a self-learning environment:

- Morphosyntactic taggers and syntactic parsers (based on some partial parsing techniques, as e.g., the Constraint Grammar theory) will provide meaningful analysis of the written texts produced by the user.
- Speech-to-text converters will transcribe the oral production and the taggers and parsers will thus be able to evaluate this oral production.
- Statistical tools will contribute to the overall assessment of the user performance by matching user production with domain-specific grammatical structures and vocabulary.
- Information extraction tools will establish the domain vocabulary and the lexical and conceptual relations that hold in the domain texts. This tool will perform a mapping of input text into a concept-based network structure along the lines of the ones developed by Wordnet, Eurowordnet, and CYC.

2.3 The Extensive Development of Linguistic Resources

In order to design a successful CALL system, the availability of adequate linguistic resources for exercise development and support material is critical. In ALLES, these resources will be

gathered according to a clearly defined methodology which differentiates online sources from textual corpora on the business and economy domain. Online sources like specialized dictionaries, grammatical repositories, terminology managers, newspapers, magazines, reports, etc. will be incorporated as materials in the support and learning tools, whereas textual corpora will be used to define statistical measures on idioms, collocations, lexical units, grammatical structures, and conceptual structures. Textual corpora will also be very helpful when designing exercises for the listening and reading comprehension tools.

2.4 The Design and Development of an Advanced System

An advanced system for long-distance language education requires a careful consideration on a software architecture that can integrate the different learning and support modules seamlessly. E-learning environments also place a special emphasis on remote access through Internet as well as on good graphical user interfaces that should be user-friendly, intuitive and powerful enough to integrate the latest web-based multimedia technologies. In ALLES, the interaction among the learning and support modules will be carefully studied according to a didactic design that can ensure an effective learning path for the user. The resulting system should also be flexible enough to guarantee continuous long-distance learning and to allow for customization by third parties. Rigorous testing environments and deployment resources for ALLES will be provided by the participants, given their extensive experience in language education.