

Evaluating integrated NLP in foreign language learning: technology meets pedagogy

Irena Vitanova

Dept. of Management, Sofia University
Tsarigradsko shose, bl. 113V, 1113 Sofia, Bulgaria
itv@gmx.co.uk

Abstract

In this paper I would like to present the pedagogical perspective upon a 3-year research project with direct implications for the field of Foreign Language Teaching (FLT), and English for Professional Purposes (ESP) in particular.

As a project in Computer Science, the basic aim of LARFLAST¹ was to investigate the integration of several innovative software components into a harmonized environment for learning English financial terminology. In order to ensure the pedagogical relevance of the work, user involvement and endorsement of the research prototype were sought at every stage of the project.

The working bench consisted of the following resources: 1) a training corpus of texts for self-study 2) a collection of additional relevant texts, gathered by an intelligent Web-agent 3) a terminology testing section of exercises in the popular CALL formats, including a NLP component 4) online financial dictionaries 5) GenWeb, generating dynamic web pages, intended as immersive contexts for the learners, making use of the capabilities provided by web browsers.

LARFLAST studied and integrated language technologies like semantic analysis, information retrieval and elements of natural language generation.

The paper discusses the difficult negotiation process between user requirements, pedagogical rationale and the specificity of technological problems that participants sought to solve and subsequent work, trying to cope with some of the problems.

1 Introduction

No matter how diversified a portfolio of methods a language teacher has, practice shows that the traditional classroom is willing to transform itself into a hybrid environment. Face-to-face communication is often complemented with CALL

formats, be they Web-based or not, intelligent or not, depending on the specific pedagogic task and technological givens of the institution.

However, good innovation design and implementation are crucially dependent on evaluation.

From a pedagogic point of view CALL provides fertile ground for innovative and highly creative work in foreign language teaching and research. As an inherently multidisciplinary endeavour, it is by necessity approached and estimated from different standpoints. An ingenious technological solution may not always find its educational application right away. On the other hand teachers, FLT researchers and learners may find substantial value in simpler and more modest applications which unleash their creativity, flexibility and autonomy.

In the following two sections I will make a brief overview of the two groups of CALL applications in view of their acceptance by the teaching expert community. I will also suggest that these evaluations represent the experience of professionals from technologically developed societies. Using the Bulgarian example I will argue that the restrictions of technological support for educational purposes encourages the teaching community to adopt and adapt ICALL solutions and make the most of them, thus giving new life to centuries-old traditions in language learning. Section 4 describes the pedagogical support for the prototype, based on an initial user-study. Section 5 deals with user evaluation and pedagogic issues arising from it. Section 6 contains the conclusion.

2. CALL without language technology

For the time being non-language technology seems to be better accepted by the language-teaching community in technologically advanced countries. Common CALL applications are automated language exercises, the use of hyper text for access to explanatory material, digital audio and video, database technology to record and present student work, as well as network communication (Nerbonne 2002, Levy 1997). The FLT community has made extensive use of eLearning platforms like WebCT [1], CISCO [2] and the freeware HotPotatoes [3] which represent

¹ LARFLAST (Learning Foreign Language Scientific Terminology) Copernicus'98 Joint Research Project (1998-2001).

empty shells where the teacher uploads training materials by means of specialized authoring tools.

Non-language technology applications in CALL generally meet enthusiastic reception – and for good reason. On the one hand, they allow for pedagogical innovation and present opportunities for those teachers who are willing to keep in touch with current technology. On the other hand, students gain unprecedented language practice through internet-based, multi-use, interactive environments. Apart from these practical effects, it provides fertile ground for the research on technology-based learning and language acquisition.

And yet there is a small, but significant snag in the usage of these applications. They presuppose the existence of widely accessible computer facilities for learners, which is still not the most common scenario for countries like Bulgaria.

In summary and transition to other applications let me quote a somewhat ironic comment: “The CALL ‘killer apps’ have been e-mail, chat and multimedia programs, developed and used by language teaching professionals with very little input from language technology research.” (Borin 2002)

3. CALL with language technology

In his overview of NLP from the CALL perspective John Nerbonne (Nerbonne 2002) states: “There are voices which urge caution in applying NLP to the CALL problem. (Salaberry 1996) assesses the suitability of language technology for CALL quite negatively, but not atypically ... attributing it to NLP’s failure to produce full-fledged intelligent dialogue in unrestricted form”.

On the other hand, those NLP contributions which have been enlisted in CALL have proved a very powerful tool in the hands of experienced teachers and imaginative learners. What is more, they have leased new life to traditional, time-tested methods of teaching and learning. The first three examples enumerated below are of technologies which are currently (and very successfully) integrated in language teaching at Sofia University.

Language corpora provide access to authentic language use to advanced and intermediate learners. When supplemented with concordance programmes, corpora are employed in a number of ways and for a number of tasks. By shifting the focus on observation, description and hypothesis formulation on the part of learners, the teacher engages them in research activity, sometimes described as “data driven learning” (Johns and King 1991).

Bilingual corpora provide not only authentic examples of target language, but comparison to mother tongue as well. The comparison may focus on specific lexical or grammatical items. Authentic bilingual corpora have been used to highlight genre distinctions in Bulgarian and English for the purposes of business communication.

Alignment techniques for bilingual texts (Paskaleva and Mihov 1998) are a tool which saves a lot of effort on the part of the lecturer, but also stimulates analysis by the learner.

In all three cases these tools are used as a complement to standard classroom techniques. In most cases students have to work independently with the corpus upon a predefined task. Occasionally the lecturer provides print-outs of relevant pages and sets an assignment. The tasks stimulate creativity and are generally more welcome than traditional philological material. Further on, **Morphological processing and generation** are used to provide exercise material for teachers, to automate dictionary access, to explain the grammatical component of morphemes.

Error recognition and diagnostics present a challenging problematic for NLP practitioners, and are at the same time charged with high expectations on the part of learners and teachers. Making errors, correcting or *not* correcting them, pointing to their source make up the reality of a real-life classroom and the subject of research in FLT. Within CALL however, (Tschichold 1999) compared automatic grammar checkers to the practice of language teachers, concluding that so far they only have very limited use.

Recently, Latent Semantic Analysis (LSA) has been used in various experiments on assessing student essays, e.g. (Foltz 1999a, Foltz 1999b).

It is within the pedagogic framework of error recognition and diagnostics that the work presented in this paper is to be placed. The general assumption is that the system diagnoses mistakes and stimulates the analysis of their source, rather than estimates and marks the student on its own.

4 Initial pedagogic support for STyLE (Scientific Terminology Learning Environment)

The development of STyLE was preceded by a user-study (Vitanova 1999) which emphasized the significance of foregrounding terminology when teaching foreign languages for professional purposes, in contrast to traditional approaches to the problem. Two relevant target groups were indicated – university students of Economics and Business Administration, and professionals in the field of Finance. The basic assumption underlying the pedagogic rationale of STyLE relies on the practical experience of numerous teachers of

foreign languages for professional purposes and can be briefly summarized as follows: Domain knowledge (in this case knowledge of the financial domain) has an effect on the acquisition of linguistic knowledge. While university students are in the process of acquiring domain knowledge, specialists can use it as a more or less firm foundation on which to base emerging linguistic structures. Hence, some linguistic mistakes can be attributed to insufficient domain knowledge, and it is here that STyLE tries to help.

4.1. Typical mistakes with respect to terminology

A series of tests with the two target groups indicated several possible **sources** of errors in user answers:

- Insufficient linguistic knowledge (leading to grammatical mistakes)
- Misunderstanding of question posed by teacher
- Lack of knowledge of correct term in foreign language
- Meaning of concept is familiar, but its name in the foreign language is wrong
- Insufficient knowledge of a sector of the domain

4.2. Setting the pedagogic task

Within the self-tuition workbench of STyLE the student has to read a short passage on a given topic, e.g. *Financial markets*, which belongs to the academic expository genre, and presents the target terms and domain knowledge in a manner corresponding to an intermediate level of proficiency of the users. Then the student has to do a series of exercises testing his reading comprehension of the text. In one of the exercises the user is asked questions on the contents of the texts he has read and has to type in freely his answer. The first version of the NLU component (employing deep understanding of correct sentences and proving the domain correctness and the appropriateness of the logical form of each sentence) required a syntactically complete and correct answer. Within the more recent version (employing shallow parsing and elements of deep semantic analysis) the student is not so restricted – he can put in phrases, loosely connected and not necessarily syntactically correct.

5. Diagnostics and feedback - general requirements from a pedagogic point of view

At this point let me remind what particular pedagogical tasks are being set and solved in this interaction.

a. The most general is that of checking the reading comprehension of a particular text within a particular genre.

b. The more specific is that of establishing what elements of the text the user does not understand.

c. And the third – to find out why this infelicity occurs – whether because the user does not know specific terms, or relational terms, or cohesive elements preventing him/her from making out the meaning, or finally – whether the learner has insufficient domain knowledge.

In traditional face-to-face communication the teacher has a multiplicity of techniques to elicit the sources of errors.

In self-tuition however, the user is left with what the system offers as feedback, hence the appropriateness of feedback is crucial.

5.1. Feedback in NLU-component of STyLE

The system offered the following set of diagnostics for freely-typed answers to questions:

Correct; Wrong; Incomplete; More specific; Paraphrase, More general or combinations thereof, e.g. **“partially correct, more general”**.

Feedback terms are not followed by any elaboration on what makes the answer “more general” or “more specific”, or even “wrong”. Thus, users’ answers are classified according to some criterion which is known to the system, but not to its users. The teachers would also have a difficulty in interpreting this feedback. They would normally proceed to correcting a mistake and identifying its source without necessarily verbalizing an interim term like “Your answer is too specific”.

5.2. User reactions and unexpected results

All seven users who tested the system found the feedback on their answers insufficient and puzzling. They would welcome as much feedback as possible, on all types of mistakes or imprecisions. One of the users (who happened to be a computer specialist as well) wrote the following comment: *“Below are my answers and the system’s response, I felt I didn’t expect these responses from the system, you can check if they are correct. I assume the word ‘market’ is recognized, so that the answer is incomplete...”*

During the discussion of system feedback users wished to be explained the criteria according to which some answers were judged “incomplete” or “too specific”. Therefore, some system readings were made available, so the learners saw the propositions comprising the minimum and maximum answers, as well as samples of learners utterances, corresponding to the types of feedback terms. Thus, if the question was *What is a primary*

market? and the answer was *The primary market is a financial market that operates with newly issued debt instruments and securities for instant delivery*, the feedback would be *Partially correct* answer, because one of the concepts used – *for instant delivery* – is wrong.

This triggered a valuable discussion which revolved around the following issues (Vitanova 2001):

- a. What is a correct/incorrect definition in the field of Finance (compared to that of mathematics, for instance)?
- b. What is a correct/incorrect definition of a financial term in an English language class?
- c. What counts as a mistake in either field?
- d. How do teaching experts react to mistakes? To what purpose?
- e. What do ESP teachers teach – language, domain knowledge, or both? In what proportions?

6 Conclusion: If this had happened in real life ...

In a traditional classroom exercises for reading comprehension accompany practically every lesson. The good thing about them is that they simultaneously provide an opportunity to practice speaking skills. In answering a question students make mistakes, but in is accepted practice that teachers do not correct all of them, so as not to interrupt a spontaneous speaking effort and create a lasting inhibition. Certain mistakes, however, are all too welcome. To use the above example of “primary market”, *any* answer – no matter if incomplete or partially wrong – will encourage the teacher to elicit further information by asking additional questions, making students analyse certain answers, and encourage a discussion until all the points deemed important are made clear. In the course of this communication classifications of answers as “wrong” or “too general” are rarely made, they are only implied in the course of the discussion.

This is not to claim that system-user interaction in NLP applications should try to reproduce teaching strategies. It is worth remembering though, that the user interacts with the system only at interface level and within the framework of the educational goals s/he has set himself. What is more, it is preferable if feedback occurs in a format which is either already familiar, or intuitively understood.

7 Acknowledgements

I am greatly indebted to Prof. Galya Angelova from the Bulgarian Academy of Sciences who was patient enough to introduce me to the

“underworld” of CALL and who is a remarkable team-builder.

References

- [1] WebCT, <http://www.webct.com/>
- [2] CISCO, <http://cisco.netacad.net/>
- [3] HotPotatoes, <http://web.uvic.ca/hrd/hotpot/>
- L. Borin. 2002. *Where will the Standards for Intelligent Computer-Assisted Language Learning Come from?* In Workshop Proceedigs. International Conference on Terminology and Language Resources Management. Las Palmas, Spain: ELRA. 2002. 61-68
- S. Jager, J. Nerbonne and A. van Essen, eds. 1998. *Language Teaching and Language Technology*. Lisse: Swets and Zeitlinger.
- J. Nerbonne 2002. *Computer-Assisted Language Learning and Natural Language Processing*. In: Ruslan Mitkov (ed.) *Handbook of Computational Linguistics*, Oxford Univ. Press, pp. 670-698.
- E. Paskaleva and S. Mihov. 1998. *Second Language Acquisition from Aligned Corpora*. In “Jager, 1998. *Language Teaching and Language Technology*. S., J. Nerbonne and A. van Essen, eds. Pp 43-52 Lisse: Swets and Zeitlinger.
- C. Tschigold, 1999. *Grammar Checkers for CALL: Strategies for Improving Foreign Language Grammar Checkers*. In: “CALL: Media, Design and Applications” K.Cameron, ed. Lisse: Swets and Zeitlinger.
- G. Angelova, S. Boytcheva, O. Kalaydjiev, S. Trausan-Matu, P. Nakov and A. Strupchanska. 2002. *Adaptivity in Web-based CALL*, Proceedings of ECAI – 2002, 445-449.
- I. Vitanova, 1999, *Learning Foreign Language Terminology: The User Perspective*. Larflast report 8.1., delivered to EC.
- I. Vitanova, 2001, *Evaluation of User Interaction with STyLE-2000 Prototype*, Larflast final report 8.1. delivered to the EC.
- P. W. Foltz, Laham D. and Landauer T. K. 1999a. *Automated Essay Scoring: Applications to Educational Technology*, In “Proceedings of the ED-MEDIA Conference, Seattle.
- P. W. Foltz, Laham D. and Landauer T. K. 1999b. *The Intelligent Essay Assessor: Applications to Educational Technology*, Interactive Multimedia Elelctronic Journal of Computer-Enhanced Learning, 1(2).