ReCALL **23**(1): 4–24. 2011 © European Association for Computer Assisted Language Learning
doi:10.1017/S0958344010000261

On using intelligent computer-assisted language learning in real-life foreign language teaching and learning

LUIZ A. AMARAL
Department of Languages, Literatures and Cultures, University of Massachusetts, Amherst, MA, USA
(email: amaral@spanport.umass.edu)

DETMAR MEURERS
Seminar für Sprachwissenschaft, Universität Tübingen, Germany
(email: dm@sfs.uni-tuebingen.de)

Abstract

This paper explores the motivation and prerequisites for successful integration of Intelligent Computer-Assisted Language Learning (ICALL) tools into current foreign language teaching and learning (FLTL) practice. We focus on two aspects, which we argue to be important for effective ICALL system development and use: (i) the relationship between activity design and restrictions needed to make natural language processing tractable and reliable, and (ii) pedagogical considerations and the influence of activity design choices on the integration of ICALL systems into FLTL practice.

Keywords: ICALL, Intelligent Tutoring Systems, activity design, NLP

1 Introduction

Computers are widely used in foreign language teaching and learning (FLTL) to help learners experience the target language and culture. Currently available technology, for example, supports multimedia presentations, web-based distribution of print media, radio, and TV programs, as well as various forms of computer-mediated communication with native speakers. Such tools help fulfill the undisputed need for contextualized, communicative language use in the acquisition process.

At the same time, an important strand of research since the 1980s (cf., e.g., Long, 1988, 1991, 1996; Ellis, 1994; Schmidt, 1995; Lyster, 1998; Lightbown & Spada, 1999; Norris & Ortega, 2000) has shown that awareness of language forms and rules is important for an adult to successfully acquire a foreign language. Schmidt (1995) argues for the importance of noticing linguistic categories and features for learning to take place. Moreover, for more than twenty years, researchers in CALL have been arguing about the importance of taking into consideration the research in second language acquisition and current foreign language teaching practice when designing CALL applications (cf., e.g., Barrutia, 1985; Levy, 1997).

In a series of interviews with Spanish and Portuguese instructors at The Ohio State University, we found that instructors perceive the inability of students to handle appropriate linguistic forms as a main obstacle in reaching the communicative goals of meaning-based activities (Amaral, 2011). On the other hand, the same instructors perceive form-based activities as problematic for use in the classroom because they can reduce the pace of the lesson and take away time that could be dedicated to meaning-based, communicative activities. In such a setting, the amount of time a student spends in a class with a language instructor is very limited, and individual interaction between instructor and student even more scarce. The consequence is that classroom time is often used for meaning-based activities, and work on linguistic categories and rules is de-emphasized and confined to homework.

When asked about the possible roles of computer tools, instructors were skeptical about computers replacing human interaction in FLTL. They pointed out that activities that involve discussion, role play, sharing opinions, and seeing language fulfill a communicative purpose in general presented ideal scenarios for students to interact with another human being. They argued that learning to communicate in a foreign language implies learning to negotiate meaning, understanding social behavior, and observing different body language strategies. The instructors interviewed were skeptical about the possibility of computers replacing humans in this respect. However, they were very receptive to the idea of automatic support tools to practice receptive skills, reinforce the acquisition of language forms, propose remedial work, and raise linguistic awareness in general.

This situation can be seen as an excellent motivation and opportunity for developing CALL tools that provide personalized feedback on learner errors and foster linguistic awareness of relevant language forms and categories. But traditional CALL systems provide only limited exercise types, such as multiple choice, matching, point-and-click, or simple form filling. Moreover, their ability to provide feedback is restricted to letter-by-letter feedback or specific hand-specified feedback messages based on matching the learner response with a pre-stored target answer, either directly or by using regular expressions. There is thus a clear need for linguistic modeling to improve the ability of CALL systems to handle more complex exercise types, and to provide detailed individualized feedback.

Research in Intelligent Computer-Assisted Language Learning (ICALL) has explored techniques and tools from Natural Language Processing (NLP) for this purpose, such as tokenizers, morphological analyzers, part-of-speech taggers, chunkers, parsers, or semantic analysis tools. In the last thirty years, a significant number of research projects have focused on topics in ICALL (cf., Engel, Bouwhuis, Bosser & D'Ydewalle, 1992; Swartz & Yazdani, 1992; Chanier, 1994; Holland, Kaplan & Sams, 1995; Heift & Schulze, 2007). Nevertheless, apart from the system presented in this paper, there are only two systems that use NLP technology and are fully integrated into real-life foreign language programs in universities: Robo-Sensei (Nagata, 2002) and E-Tutor (Heift, 1998, 2003).[1]

[1] See the CALICO software reviews, accessible from https://calico.org, for a comprehensive list of CALL programs. Generally, those programs either do not use NLP to process learner

Why are CALL systems that use NLP technology not more commonly developed and used, despite their clear potential to contribute to FLTL? Why are there so many projects that explore NLP technology in CALL, yet so few result in tools that are used by actual language learners? How can NLP technology be effectively employed to create ICALL systems that are useful and usable in real life?

In this paper, we review the recent history of ICALL and identify some major challenges ICALL researchers and system designers face. We then describe a concrete way of addressing these challenges and illustrate it based on an ICALL system we developed for the instruction of Portuguese. The focus of our discussion is on systems that are designed to be integrated into FLTL curricula. In most cases, these systems present activities that are incorporated into everyday FLTL practice. Because we are interested in systems that fulfill specific pedagogical goals, we exclude systems from our analysis that use NLP but function more like a writer's aid, such as ICICLE (Michaud, McCoy & Pennington, 2000) and El Corrector (http://www.translation.net/el_corrector.html).

The overall goal of this paper is to bring two aspects into the debate which we believe are important for effective ICALL system development: (i) the relationship between activity design and restrictions needed to make NLP analysis tractable and reliable, and (ii) pedagogical considerations and the influence of activity design choices on the integration of ICALL systems into FLTL practice.

2 ICALL today

Garrett (1995) identified the integration of foreign language teaching expertise into the development of ICALL systems as one of the main challenges of ICALL research and development. She also criticized the lack of support for teacher involvement with technology in post-secondary education. Today, the latter no longer appears to be as significant a problem. Most language departments in North America encourage the use of technology in foreign language teaching, the number of positions for CALL experts in such departments has grown, and CALL conferences such as CALICO and EUROCALL are attended by participants from all over the world. At the same time, the development of systems using NLP technology is not on the agenda of most CALL experts, and interdisciplinary research projects integrating computational linguists and foreign language teachers remain very rare.

To determine the potential role of ICALL systems in FLTL, arguably it is vital to understand the instructional methods and approaches used in FLTL and to establish the needs of students and teachers. While some general studies of the expectations of such professionals exist (Levy, 1997; Atwell, 1998, 1999), there is little discussion about the integration of ICALL into the methodologies currently used in FLTL.

On the computational side, much has been written in the last twenty years about the development of technologies that allow computers to deal with learner language (cf., e.g., Holland *et al.*, 1995; L'Haire & Faltin, 2003; Heift & Schulze,

(F'note continued)
input and provide feedback, or they are writer's aid type programs that do not propose specific FLTL activities. In the second case, reviewers typically complain about the reliability of the feedback provided.

2003, 2007; Dodigovic, 2005). Different approaches to parsing erroneous input have been proposed, with defenders of both constraint relaxation techniques (cf., e.g., Schwind, 1995; Vandeventer, 2001; Reuer, 2003) and malrule techniques (cf., e.g., Weischedel, Voge & James, 1978; Sleeman, 1982; Covington & Weinrich, 1991) arguing about the best way to identify errors in student sentences. The focus of these research projects typically was on the development of algorithms applied to hand-constructed examples. Just as in other areas of NLP in the last century, evaluation with authentic data received only little attention; some early attempts using small learner corpora for evaluation are discussed by Heift and Schulze (2007: 59).

In addition to this empirical disconnect, there also is a deficit in ICALL research concerning the context in which learner language is produced. Being able to process ill-formed input is only part of the challenge of designing real-life ICALL systems. Issues such as activity design, language assessment and measurement, teaching techniques, syllabus design, second language analysis, cognitive models of second language acquisition, and language policy and planning are important for the design of ICALL systems for real-life FLTL. Yet these are issues outside the area of expertise of computational linguists and computer scientists behind most ICALL projects. The lack of interdisciplinary research combining NLP, SLA and FLTL expertise thus can be identified as one of the main reasons for the very limited impact of ICALL on foreign language programs.

Fortunately, there are also some exceptions, showcasing successful integration of NLP technology in ICALL systems that are fully integrated into language programs. *Robo-Sensei* (Nagata, 2002), and *E-Tutor* (Heift, 1998, 2003) are two such successful examples. There is also a third one, *Spanish for Business Professionals* (Hagen, 1999), although in this case the system was not designed to be integrated into a language program, but rather to be used as a stand-alone product.

In section 2.1, we briefly describe these three systems, focusing on activity design and the nature of the learner input to the system, the use of L1 by the system, and the use of learner modeling for adapting feedback messages or the sequencing of instruction. On this basis, in section 3 we characterize the challenges ICALL designers face in terms of how to obtain reliably processable learner input for activities that are in accordance with current FLTL methodologies. In section 4, we then discuss how we decided to address those challenges in our own ICALL system development.

2.1 Current systems

2.1.1 Robo-Sensei (Nagata, 2002, 2009) The Robo-Sensei system for Japanese presents a series of exercises for each of its 24 lessons. The activities are nicely contextualized, and there are visual aids for each, with pictures of Japan or Japanese drawings. It explicitly cross-references current textbooks, so that instructors can choose activities according to classroom material. Nagata has done extensive research on the use and effectiveness of the system (Nagata, 1993, 1996, 1997).

Constraining input and activity design: A characteristic of Robo-Sensei's activity types which is relevant in the context of this paper is the extensive use of translation to elicit student answers: English cues are given frequently to control what students enter into the system in Japanese so that, in effect, such activities can be seen as contextualized translation exercises.

Use of L1: A related challenge that systems such as Robo-Sensei have to deal with is the question how much L1 is desirable in its activities and feedback messages. For systems that deal with languages that use non-Roman alphabets, deciding when to use L1 is particularly complex. In Robo-Sensei, the choice is to convey all information related to activity descriptions and exercise instructions in English. The feedback messages generated by the system are also in English, and some of the answer triggers and examples use language comparisons between English and Japanese.

Learner modeling: In the Robo-Sensei system, the same sequence of activities is presented to every student and for a given learner input the system always provides the same feedback. The system thus is not user-adaptive in the sense that it does not take into consideration a student's levels of proficiency, knowledge about language items, or ability to deal with different types of activities when providing feedback.

2.1.2 Spanish for Business Professionals (Hagen, 1999) Spanish for Business Professionals (SBP) is a program to teach business Spanish in twelve units. The units are well-contextualized with an excellent selection of audio material. There is an interesting progression of exercises and one finds good visual aids. The program also presents several help tools, including links to grammar explanations, and words in the texts are linked to an electronic bilingual dictionary.

Constraining input and activity design: SBP is another example of how delicate is the issue of eliciting student input in ICALL. Some of its exercises make extensive use of English material to be translated by the learner to guarantee that the input provided by the learner can be processed by the system. Its vocabulary exercises, for example, are based on simple word translations, which makes them look like old-fashioned drill activities whose triggers are L1 lexical items. Another way of requesting student input that is used by the system is dictation. Decontextualized dictation is used to elicit more complex structures in more advanced activities. An interesting exercise offered by the system is called *Charadas*. It requires students to unscramble letters in a word, and then place words in the correct order in a sentence. Although it presents an interesting way to elicit students' answers, it is still not meaning-based. Both *Charadas* and the translation exercises use NLP to provide feedback to students. The reading comprehension activities offered by SBP are limited to questions with multiple choice answers. Where meaning plays a role in an activity, the system thus constrains the learner input in a way eliminating the need for NLP analysis.

Use of L1: Most of the SBP interface is in English, although the menu for the lessons and the links to online news are in Spanish as the target language. The grammar explanations provided with the units are also in English. For exercises that use NLP, the L1 is used as a trigger for vocabulary and grammar activities, which makes many of its activities look like translation exercises.

Learner modeling: Just as Robo-Sensei, SBP is not user-adaptive. Its flow of instruction is predefined and feedback is provided based solely on the type of error detected by the system in the given input. SBP uses grammatical terminology in L1 to provide feedback, independent of the learner's needs and abilities beyond those exhibited by a single learner response.

2.1.3 E-tutor (Heift, 2003) The E-tutor is an ICALL system developed by Trude Heift. It is fully incorporated into the German curriculum at Simon Fraser University, where German students enrolled in regular classes complete E-tutor exercises as a regular requirement. The E-tutor (originally called German Tutor) includes a modern web interface and the system has evolved significantly over the years, with a number of publications documenting the system development and effectiveness (Heift, 2001, 2004, 2005).

Constraining input and activity design: The system includes four types of exercises where NLP components are used to generate feedback: provide the missing word(s), build sentences with the words given, translate a phrase, and write down the sentence read by the system.

Parallel to the strategies used by the two systems discussed above, the elicitation techniques used to constrain the student input for activities with automatic analysis of learner language are thus translation, dictation, and lists of words.

Use of L1: The E-Tutor uses English as the primary language of instruction. Although instructions are given in English, E-Tutor activities such as 'build a sentence', 'provide the missing word' and 'listen to a sentence' use the target language, German, to elicit the input so that, different from other systems, these activities do not involve translation. For feedback messages, positive short feedback is provided in the target language whereas longer explanations about error types are in English.

Learner modeling: The system incorporates an explicit learner model. This allows the system to generate reports about a student's performance, a very useful tool for instructors and students. Feedback messages are also based on the student's level of proficiency in the specific grammar skill related to the error. Beginner students see the error highlighted and a grammar explanation in English, intermediate students see the error highlighted and a classification for that type of error, and an advanced student only sees the error highlighted (Heift, 2005).

3 Challenges in developing ICALL systems

The three ICALL systems sketched above can only be effective because the student input elicited by their activities can be processed by their NLP components. To achieve this, the system designers had to consider the processing capabilities and decide on the types of exercises that are suitable for what their NLP modules can handle. This is where interdisciplinary collaboration is vital, given that decisions made during this design step directly affect the usefulness and acceptability of the resulting system for foreign language learners and instructors.

In this section, we discuss four challenges for ICALL system design, from the processing capability of systems to foreign language teaching methodology. The list of challenges is not meant to be exhaustive, but represents important issues ICALL development has to face.

3.1 Challenge 1: Constraining the learner input to the system

The first challenge concerns the question how the learner input can be constrained so that the expected student input to the system can be processed effectively and efficiently.

We first motivate why it is important to consider how the system input can be characterized and constrained and then discuss some of the consequences for system design.

3.1.1 Why constrain the learner input? A key aspect in the development of efficient ICALL systems is the reliability of its feedback messages. Research has shown that, different from writer's aid tools for native speakers, ICALL systems have to provide feedback with very high precision (cf., e.g., Tschichold, 1999, 2003). Most grammar checkers for native speakers rely on the user's evaluation of its analysis to confirm if the modifications suggested by the system fit the text passage in question in terms of grammatical accuracy and style. ICALL systems for language learners, on the other hand, are expected to behave as experts that provide unequivocal analysis and precise feedback to the learner. Language learners are generally not able to evaluate the system feedback in the way a native speaker is expected to. Correspondingly, it is better for an ICALL system to provide no analysis at all to a given input than to overflag and provide incorrect feedback by reporting an error for a well-formed passage. Yet, for a number of reasons it is particularly difficult for NLP to analyze learner language and provide feedback with high precision.

Search space of syntactic processing: The first difficulty for an NLP analysis of learner language results from the enormous search space that can easily arise in the recursive morpho-syntactic processing of learner language. The lexical and syntactic properties of a language are normally used to define and constrain the search space of syntactic parsing. But just by looking at a given learner sentence, it is unclear which of the regularities of the language to be acquired are actually followed by a learner. After all, in certain contexts, "I will not buy this record, it is scratched" can even be a foreigner's request to buy cigarettes! In the *Dirty Hungarian Phrase Book* sketch by Monty Python that this example is taken from, the context that supports this deviation from the standard, native English analysis is the use of an incompetent phrase book. While this is an extreme case, it nicely highlights that the interpretation of a learner utterance will often require additional information about the context, in order to be analyzable in a way supporting meaningful feedback.[2]

Making matters worse, in order to successfully analyze syntactic structure, a parser typically has to handle structural ambiguities, which are common in language. The parser has to deal with an even greater number of possible ambiguities when processing ill-formed input, given that it has to generate all possible parses using either a greater number of rules (e.g., augmenting the grammar with *mal*-rules) or limiting the scope of restrictions which are part of the native rule set (e.g., relaxation of constraints enforcing agreement). Lexical ambiguity and incorrect or unusual word choice by foreign language learners can complicate matters further. Without additional restrictions on the nature of the *ill-formed variation* to be expected for particular contexts and learners, the number of possible combinations of rules to parse a learner sentence thus can easily become intractable.

[2] The sketch also contains an instructive example of feedback failing to produce the desired effect, when the shopkeeper's reply, "No, no, no. This is a tobacconist's" results in the correction, "I will not buy this tobacconist's, it is scratched." See http://purl.org/net/mp-sketch

Evaluating meaning: The second difficulty for NLP is related to meaning and how it can be analyzed. Human languages generally make it possible to express the same meaning in a number of different ways, including differences in how the overall meaning to be expressed is parceled out into linguistically realizable units, and the wide range of syntactic constructions and lexical choices that are available. Thus, for an ICALL system it is at least as difficult to determine if an answer is correct in terms of meaning (in order to determine whether it successfully completes a meaning-based activity) as it is to determine if it is wrong in terms of form (in order to provide feedback on form errors).

Most NLP technology used in current ICALL systems is designed to deal with morphosyntactic processing. There are checks for spelling, agreement, subcategorization, and similar types of errors that spell-checkers, morphological analyzers or parsers can identify. Meaning based processing, on the other hand, has received much less attention in ICALL so far (but see Ramsay & Mirzaiean, 2005; Delmonte, 2003; Bailey & Meurers, 2008). ICALL systems thus generally compare the meaning of a learner to the meaning of a target answer by matching surface representations such as strings or tokens. Evaluating meaning by comparing surface forms is only possible when a given meaning has to be expressed in a particular way. Yet, as mentioned above, human language typically offers a number of different ways to express a given meaning, so that one needs to expect significant *well-formed variation* in the learner input to an ICALL system unless additional restrictions on the input are imposed.

In sum, in order to obtain tractable and reliable NLP supporting the analysis of both form and meaning, it is necessary to restrict the ill-formed and well-formed variation in learner input that an ICALL system needs to deal with.[3]

3.1.2 How can the input be constrained? The question of how the input can be constrained is directly related to what triggers the learner input to the system. In the context of a web-based workbook, this is the activity that the learner is providing the input for. The most direct way of constraining the input is to explicitly require the learner to use certain language material, e.g., by specifying that certain L2 words must be used in the answer or by providing an L1 sentence to be translated by the learner. We saw in section 2.1 that all current ICALL systems make use of such explicit requirements in the design of the activities they offer.

On the other hand, it is attractive to investigate how the input can be constrained implicitly in order to provide more space for negotiation of meaning as needed for meaning-based activities. In section 4, we thus explore how an ICALL system can present sound meaning-based activities that sufficiently constrain the learner input to obtain tractable and reliable NLP analysis.

[3] Confirming our analysis in practice, projects that in recent years have aimed at processing unconstrained input have faced serious problems. El Corrector was severely criticized for its unreliability (Klein, 1998) and for the very ambitious FreeText project one reads: "Par rapport à nos ambitions de départ, FreeText a vu ses ambitions réduites. La technique de la comparaison de phrases a dû être reportée après la fin du projet et les performances du système de diagnostic peuvent sembler relativement faibles par rapport aux espoirs soulevés." (L'Haire, 2004: 10).

3.2 Challenge 2: Activity specification and instructions

The second challenge for ICALL system development to be discussed here is related to activity specification and instructions for activities. As we saw in the previous section, activity design plays an important role in constraining student input. This brings us to the question of how such activities are to be presented to the learner. A student may be asked to perform some activities that require more than one simple step. However, the more complex the activity is, the more precise and complete the instructions have to be in order to guarantee that the expectable variation in the learner input remains within the reach of what can be reliably processed. It is thus challenging to provide instruction in a concise and effective way that at the same time is pedagogically sound.

Some ICALL systems, such as the SBP system we introduced in section 2.1.2, include extensive written instructions for the exercises. As discussed by Hémard (2004: 513), texts that require students to scroll down through long pages, and are not interactive in nature, result in less productive interaction on the part of the learner. We will see in section 4.3 that careful interface design can significantly reduce the need for long exercise instructions.

In SBP, exercise instructions sometimes also occur mixed in with system instructions about how to operate the system and with language instructions about the linguistic structures. Without going further into this issue in this paper, we would argue for separating the three conceptually distinct parts: the ICALL system and how it is used, the exercise and how to complete it, and the linguistic properties of the language to be learned.

3.3 Challenge 3: Use of L1 in activities, instructions, and feedback

A third challenge for the full integration of ICALL systems into FLTL curricula is the rather indiscriminate use of L1 often found in such systems. While current paper-based FLTL materials try to attenuate the use of L1, ICALL systems tend to rely heavily on the native language of the learner to provide instructions and feedback messages. This is a delicate matter because there is no single answer to the question of how much L1 is too much. Ultimately, ICALL designers try to avoid what one could call the "L2 halting effect" where, due to exclusive use of L2 in the activities, their instructions, or the system feedback messages, the learner neither knows what to do nor understands what the system is reporting, and in response stops working with the system. For feedback messages, it can also have the effect of preventing the learner from understanding the issue reported by the system, which severely limits the usefulness of such feedback.

There are several factors that play a role when trying to decide when to use the L1 and when the L2 in an ICALL system. The first one is the distance between L1 and L2. For example, where the L2 does not even share the same writing system with the L1, it is impossible to present all material in the target language from the very beginning. The second is the nature of the activities. More complex activities require more elaborate instructions (cf., e.g., the interaction between prompt/input data and learner response discussed by Douglas, 2000). The instructions are often formulated in the L1 in order to be comprehensible, at least for beginning learners. The third is the nature of the feedback. If designers opt for messages using grammatical terminology as the only way for providing feedback, they will feel more compelled to use the L1 to get such an abstract, analytical message across. Finally, there is the design of the system itself. If the system has a complex design with difficult navigation, written instructions in L1 may be necessary. As we will see in section 4.3, in our system design we rely instead on the page layout to make system navigation more intuitive and to reduce the dependency on L1.

3.4 Challenge 4: Feedback based on linguistic, learner and activity information

ICALL systems differentiate themselves from traditional CALL systems through their ability to analyze learner input and provide appropriate feedback. For human tutors, it is clear that providing feedback to foreign language learners on their written production involves a range of information of which the knowledge about the linguistic properties of the target language is only one part. Human tutors consider information about:

- the learner: level, age, L1, maturity, knowledge of grammatical terminology, motivation to perform the activity, etc. (cf., e.g., Dörnyei, 2005);
- the task: type of activity (reading, listening, composition writing, etc.), type of question item (*wh*-question, fill-in-the-blanks, link the columns, etc.), level of question in relation to level of student, time available, material to be consulted (dictionary, grammar book, internet), etc. (cf., e.g., Willis & Willis, 2007);
- the language: grammatical competence exhibited by the linguistic properties of the learner language (lexical, syntactic, semantic, pragmatic), the nature and type of deviations in ill-formed utterances (duplication of letters, agreement, wrong synonym, lack of anaphoric reference, etc.), level of learner language in relation to scales of language complexity and development, as well as sociolinguistic, discourse, and strategic competences (cf., e.g., Canale & Swain, 1980).

Most ICALL systems, on the other hand, only take into account the language aspect and furthermore focus exclusively on grammatical competence, for which the NLP components provide the linguistic analysis (but cf. Amaral & Meurers, 2008). On the basis of this linguistic analysis the learner typically receives a meta-linguistic message with grammatical terminology in L1 about what went wrong. The activity and context in which the error was made and individual learner differences are not taken into consideration. This contrasts not only with the human tutors mentioned above, but also differs from the research on intelligent tutoring systems in other disciplines, where student and instruction models have received significant attention and exist side by side with the expert models encoding the particular domain knowledge (here, the L2 to be acquired).

An important exception in the ICALL domain is the E-tutor (Heift, 2004) which, as mentioned in section 2.1.3, includes a student model that plays an active role in feedback selection. The ICICLE system (Michaud, McCoy & Stark, 2001), is also worth mentioning here (despite its different function as a type of writer's aid) since it includes a learner model which, in addition to choosing feedback strategies, also helps constrain the search space of parsing by selecting the rule set that is assumed to be accessible to the learner at their current stage of the acquisition process.

4 Approaching the challenges

After the general characterization of four important challenges, which arise in the development of ICALL systems for real-life language teaching, in this section we exemplify how we approached some of these challenges in the development of an ICALL system for Portuguese.

4.1 TAGARELA

TAGARELA (Teaching Aid for Grammatical Awareness, Recognition and Enhancement of Linguistic Abilities) is an ICALL system for Portuguese designed to complement existing pedagogical materials in an introduction to Portuguese at the college level. It can be seen as an electronic workbook that offers on the spot individualized feedback on spelling, morphological, syntactic and semantic errors. It provides opportunities to practice listening, reading and writing skills. TAGARELA's exercise types are similar to the ones found in regular workbooks: listening and reading comprehension, description, vocabulary practice and re-phrasing. The system was used at the Ohio State University by students in regular classroom settings as well as in an individualized instruction program. A discussion of student feedback on the system is included in Amaral (2007). The system is currently being adapted for use in a distance learning program at the University of Massachusetts.

4.2 Activity types

As a first step in designing the activities for TAGARELA we explored which activities are motivated by current FLTL methodologies and constrained enough to support effective NLP analysis, as motivated in section 3.1.

In terms of the pedagogical requirements, based on interviews with foreign language instructors (Amaral, 2011), we identified a need for activities that expose students to original listening passages, for exercises which allow students to practice reading comprehension skills, for on-the-spot feedback for activities in which students can practice writing skills, and for activities that support practice of the morphological patterns of Portuguese. We thus created six different types of activities for TAGARELA: reading comprehension, listening comprehension, picture description, rephrasing, fill-in-the-blanks, and vocabulary. The system provides immediate, individualized feedback to the learner for all activity types.

The role of meaningful interaction in the acquisition process has been much-debated in SLA (cf., e.g. Krashen, 1987, 1988; Lee & VanPatten, 1995; Leaver & Willis, 2004; Gass & Mackey, 2007). On this basis, activity types that deal exclusively with translation and dictation have received severe criticism by language teachers who adopt more meaning-based methodologies. In the same vein, activities with repetition or substitution drills, and decontextualized fill-in-the-blanks have been left out of teaching materials for their lack of communicativeness. While it is beyond the scope of this paper to argue for or against specific activity types in FLTL, when designing ICALL systems to be used in real-life foreign language programs it is important to acknowledge the fact that many language teachers and policy makers have reservations about such activities. Thus, one way of increasing the acceptability of ICALL systems is to avoid mechanisms constraining the learner input in a way that eliminates meaningful interaction.[4]

In TAGARELA, we avoid translation and dictation exercises, or any other method to restrict student input that could present problems to the FLTL methodology used. We also limit the use of L1 and the length of instructions whenever possible. At the same time, we realized the activities in a way constraining the expected input sufficiently to obtain effective NLP analyses. In the next section we discuss these design choices in more detail.

4.3 Some design choices

Visual design: While the design of the user interface has traditionally not received much attention in ICALL, the design of the activity pages does have an important impact, e.g., on the nature and type of instruction that needs to be given to the learners to use the system effectively. As Hémard (2006: 32) reminds us, "learners should be encouraged to establish a useful relationship between context and use designed to facilitate and support the understanding and communication of meanings."

In TAGARELA, each activity page has the same general page layout, which is illustrated by a listening comprehension exercise in Figure 1: on top one finds the system banner and below that the menu with the different activity types.

Below this, one finds the name of the activity and menus to select other activities of this type in other modules. The middle of the page then presents the specific activity and the instructions for it. The learner typically enters text at the bottom left of the page and the system displays feedback messages at the bottom right. The different parts of the page thus have consistent functions throughout.

Within this visual frame of reference, each of the six activity types in TAGARELA (reading, listening, description, rephrasing, fill-in-the-blank, vocabulary) uses a specific icon and page color to identify the activity type. The students thus have consistent visual cues about the nature of the activity. This is best illustrated by comparing the Listening Comprehension Exercise we saw in Figure 1 with the Reading Comprehension Exercise in Figure 2 and the Picture Description Exercise shown in Figure 3.

Language of instruction: In line with the discussion in section 3.3, we chose a hybrid approach for the language of instruction. English (L1) instructions are not avoided completely, but students always see instructions in Portuguese (L2) first. If they want to read instructions in English, they have to place their mouse point over the American flag. As soon as they move their mouse elsewhere again, the Portuguese instructions return to the screen. All activity buttons are in Portuguese, with the consistent page layout helping learners understand the function associated with each button.

[4] Note that making activities meaningful does not necessarily mean complying with authenticity criteria. Authentic activities require students to use world knowledge to produce authentic solutions for tasks that mirror the priorities and challenges of real-life situations (see Wiggins, 1998). ICALL activities have to be valid, i.e., they have to allow us to "infer real performance results for specific standards" (Wiggins, 1998: 141), but, similar to traditional workbook activities, they are almost never authentic.

Fig. 1. Listening Comprehension Exercise

General setup: An activity menu is presented to students after they log in. The menu is hierarchically structured by course level, module, and activity type. Students choose their level, module, and activity type in a buffet-style learning approach. Once a student has made a choice, the system offers the corresponding activity and provides immediate feedback whenever learner input is submitted to the system by selecting the corresponding button (*Enviar*). Similarly to other ICALL systems, TAGARELA also provides buttons with all diacritic symbols for capital and low letters. Clicking on one of those buttons inserts the accented character at the place at which the cursor is positioned in the input field.

Constraining the learner input: In TAGARELA, elicitation of learner input is never done through translation or dictation exercises. The system instead employs other

Fig. 2. Reading Comprehension Exercise

techniques to constrain student input without jeopardizing the necessary content manipulation proposed by the activities: It uses pictures, lists of words, contextualized listening passages, gap-filling, written cues in L2, or a combination of two or more of these techniques to constrain the space of potential answers that a learner might provide.

For example, in the description activity in Figure 3, the student has to describe the hotel room she sees in the picture using one of the expressions of place in the list provided with the instructions, and the words "vaso" (vase) and "mesa" (table) provided with the picture. The combination of different types of information restricts

Fig. 3. Description Exercise

the input to possible answers, which makes it possible for the system's NLP modules to reliably evaluate the semantic appropriateness of the student's answer.

The listening and reading comprehension activities as the most meaning-based activities offered by TAGARELA provide the fewest external constraints on the learner input to the system. The difficulty of the NLP needed to analyze such learner answers can, however, be limited indirectly by carefully choosing the nature of the listening passages or text, the questions asked about it, and by specifying activity models which support flexible content matching strategies of the kind discussed in Bailey and Meurers (2008). Before we turn to a discussion of the TAGARELA architecture showing how activity models, learner models, and the expert module are combined, let us point out that we view the exploration of the space of meaning-based activities and the indirect constraints which can make them computationally tractable as one of the most important areas for future work.

4.4 System architecture

Having characterized the TAGARELA system setup and the activities it offers, we turn to the question of how the system can process the student input elicited by those activities to generate appropriate feedback. Among the tasks that an intelligent language tutor can perform are (i) detect indicative learner language patterns (usually errors) in the student input, (ii) diagnose the knowledge structures and skills of the students, (iii) adapt instruction accordingly, and (iv) provide personalized feedback. Since Hartley and Sleeman (1973), an intelligent tutoring system (ITS) needed to perform these tasks is recognized as consisting of at least three components: (i) an expert model, which contains the domain knowledge (in the case of ICALL, the knowledge about language); (ii) a student model, which keeps track of a student's knowledge of the structures to be acquired; and (iii) an instruction model, which is a repository of information to achieve a better tutorial strategy. All three components are important to ensure the adequacy of feedback messages. Good linguistic modeling is necessary for processing the input, knowledge about the learner is needed to adapt feedback to each individual, and information about how to provide instruction is important to adjust feedback to the situation where the error occurred.

TAGARELA's architecture reflects the importance of combining the necessary components of an ICALL system with the requirements and specifications of the types of activities used. As we see in Figure 4, the TAGARELA architecture consists of six modules: Interface, Analysis Manager, Feedback Manager, Expert Module, Instruction Model, and Student Model.

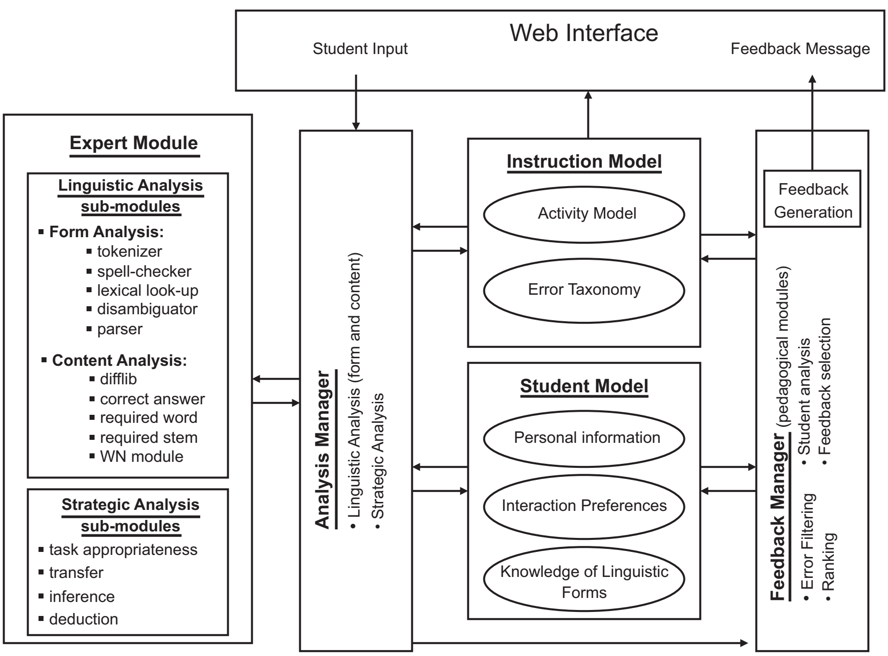

Fig. 4. TAGARELA Architecture

The *Interface* is dynamically created using the mod_python module running in an Apache2 web server. The exercise pages are generated from activity specifications, which are separate from the templates encoding the web pages. The system makes use of AJAX[5], a web programming technique that allows web clients (e.g., Firefox) to asynchronously interact in the background with the web server (in our case the Apache2 server with the mod_python module integrating the NLP). The learner can thus interact with the system and obtain feedback at any time, e.g., while the audio file for a listening comprehension activity continues to play.

The *Expert Module* is a collection of NLP sub-modules that are called in order to provide specific analysis of the input sentence. The tokenizer is designed to take into account specific properties of Portuguese, such as cliticization, contractions, and abbreviations (Amaral & Meurers, 2009). The input then is checked for non-word spelling errors with Ispell (Kuenning, 2005) as a standard spell-checker using Brazilian Portuguese parameter files. We combine lexical lookup and morphological analysis in a full-form lexical lookup step, which returns multiple lexical analyses with full morphological information from the CURUPIRA lexicon (Martins, Nunes & Hasegawa, 2003). Disambiguation rules are then used to narrow down the multiple lexical analyses based on the local context, following the general idea of Constraint Grammar (Karlsson, Voutilainen, Heikkilä & Anttila, 1995). Given that the disambiguation rules we employ are strictly local in nature, we use a simple bottom-up parser running on a small hand-written grammar to check some global well-formedness conditions, agreement, and case relations. Complementing the form-focused processing, we also perform shallow semantic matching between the learner answer and target answers provided by the instructor as part of the activity model.

The *Analysis Manager* essentially coordinates the NLP analysis of the learner answer, taking into account the processing needs specified in the activity model of a given activity. As discussed in detail in Amaral, Meurers and Ziai (2011), the learner input is annotated with the output of the NLP modules it has called and passes on the information to the Feedback Manager. The *Feedback Manager* is responsible for choosing the best feedback strategy to generate a feedback message, which is then displayed to the student. It also updates the Student Model with the information received from the Analysis Manager.

The *Instruction Model* is the component of the system responsible for storing knowledge about instructional elements that influence the system's interaction with the learner, and the *Student Model* is the repository of information about each individual student's current state of knowledge. In order to avoid false inferences about the student's linguistic competence it is important to analyze the performance of the student in relation to the type of activity where it occurs. In consequence, it is necessary to establish ways to classify activities and to provide information about them so that the system can determine which inferences may be supported by a given activity. The implemented system includes only a basic version of these models, but a detailed account and discussion of the full design can be found in Amaral and Meurers (2008).

[5] See http://en.wikipedia.org/wiki/Ajax_(programming) for a detailed characterization and related links.

The NLP architecture of TAGARELA sketched above was built to support the types of activities the system proposes. The Analysis Manager provides the flexibility necessary to analyze different types of input triggered by the different activities. For example, the processing requirements of a fill-in-the-blanks activity in which the expected input is a word differs considerably from those of a reading comprehension in which a full sentence has to be typed in. Several of the sub-modules necessary to process the latter (tokenizer, deep syntactic analysis, meaning-driven sub-modules) are dispensable for the former, which is specified in the individual activity models informing the Analysis Manager.

5 Conclusion

ICALL systems can in principle play a significant role in language learning if they are designed to address the needs of learners and instructors in the current foreign language teaching and learning context. In this paper, we identified and characterized several key challenges which ICALL researchers face when designing systems to play such a role. They range from the activity types and the methods used for constraining student input to them in order to obtain effective automatic analysis, via the interface design, instructions and the use of L1, over to the nature of the feedback messages and their basis in language analysis, learner and activity models. Making matters concrete, we discussed our approach to some of those challenges in the TAGARELA system and described the NLP architecture supporting our design choices. The TAGARELA architecture is freely available to other researchers under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.[6]

The paper makes explicit the relationship between the technology available for the implementation of ICALL systems and the consequences it can have on system design for specific pedagogical contexts. The overarching goal of the paper is to argue in favor of a multidisciplinary approach to ICALL design and implementation, which adapts current NLP approaches to the context of FLTL in order to obtain ICALL activities which are pedagogically sound and computationally tractable.

References

Amaral, L. (2007) *Designing Intelligent Language Tutoring Systems: Integrating Natural Language Processing technology into Foreign Language Teaching*. Ph.D. thesis, The Ohio State University.

Amaral, L. (2011, forthcoming) Revisiting Current Paradigms in Computer Assisted Language Learning Research and Development. *Ilha do Desterro*, **60**.

Amaral, L. and Meurers, D. (2008) From Recording Linguistic Competence to Supporting Inferences about Language Acquisition in Context: Extending the Conceptualization of Student Models for Intelligent Computer-Assisted Language Learning. *Computer-Assisted Language Learning*, **21**(4): 323–338. http://purl.org/dm/papers/amaral-meurers-call08.html.

Amaral, L. and Meurers, D. (2009) Little Things With Big Effects: On the Identification and Interpretation of Tokens for Error Diagnosis in ICALL. *CALICO Journal*, **26**(3): 580–591. http://purl.org/dm/papers/amaral-meurers-09.html.

[6] http://creativecommons.org/licenses/by-nc-sa/3.0/

Amaral, L., Meurers, D. and Ziai, R. (2011, forthcoming) Analyzing Learner Language: Towards A Flexible NLP Architecture for Intelligent Language Tutors. *Computer-Assisted Language Learning*, **24**(1). http://purl.org/dm/papers/amaral-meurers-ziai-10.html.

Atwell, E. (1998) What can SALT offer the English Teaching Professional? *English Teaching Professional*, **7**: 46–47.

Atwell, E. (1999) *The Language Machine: The Impact of Speech and Language Technologies on English Language Teaching*. London: British Council, http://www.comp.leeds.ac.uk/eric/atwell99bc.pdf.

Bailey, S. and Meurers, D. (2008) Diagnosing meaning errors in short answers to reading comprehension questions. In: Tetreault, J., Burstein, J. and De Felice, R. (eds.), *Proceedings of the 3rd Workshop on Innovative Use of NLP for Building Educational Applications, held at ACL 2008*. Columbus, Ohio: Association for Computational Linguistics, 107–115. http://aclweb.org/anthology/W08-0913.

Barrutia, R. (1985) Communicative CALL with Artificial Intelligence: some desiderata. *Computer Assisted Language Learning*, **3**(1): 37–42.

Canale, M. and Swain, M. (1980) Theoretical Basis of Communicative Approaches to Second Language Testing. *Applied Linguistics*, **1**: 1–47.

Chanier, T. (1994) Special Issue on Language Learning: Introduction. *Journal of Artificial Intelligence in Education*, **5**(4): 417–428.

Covington, M. A. and Weinrich, K. B. (1991) Unification-based Diagnosis of Language Learners' Syntax Errors. *Literary and Linguistic Computing*, **6**(3): 149–154.

Delmonte, R. (2003) Linguistic knowledge and reasoning for error diagnosis and feedback generation. *CALICO Journal*, **20**(3): 513–532.

Dodigovic, M. (2005) *Artificial Intelligence in Second Language Learning: Raising Error Awareness*. Clevedon, UK: Multilingual Matters Ltd.

Douglas, D. (2000) *Assessing Languages for Specific Purposes*. Cambridge: Cambridge University Press.

Dörnyei, Z. (2005) *The Psychology of the Language Learner: Individual Differences in Second Language Acquisition*. Mahwah, NJ: Lawrence Erlbaum Associates.

Ellis, N. (1994) Implicit and Explicit Language Learning - An Overview. In: Ellis, N. (ed.), *Implicit and Explicit Learning of Languages*. San Diego, CA: Academic Press, 1–31.

Engel, F. L., Bouwhuis, D. G., Bosser, T. and D'Ydewalle, G. (eds.) (1992) *Cognitive modelling and interactive environments in language learning*. No. 87 in NATO ASI series: Ser. F, Computer and systems sciences. Berlin: Springer. Proceedings of the NATO Advanced Research Workshop on Cognitive Modeling and Interactive Environments, held in Mierlo, The Netherlands, Nov. 5–8, 1990.

Garrett, N. (1995) ICALL and Second Language Acquisition. In: Holland, V. M., Kaplan, J. D. and Sams, M. R. (eds.) (1995) *Intelligent language tutors: Theory shaping technology*. Mahwah, NJ: Lawrence Erlbaum Associates, 345–358.

Gass, S. M. and Mackey, A. (2007) Input, interaction and output in SLA. In: Williams, J. and Patten, B. V. (eds.), *Theories in Second Language Acquisition: An introduction*. Mahwah, NJ: Lawrence Erlbaum Associates, 175–199.

Hagen, L. K. (1999) Spanish for Business Professionals. Project Web Page. http://www.uhd.edu/academic/research/sbp/.

Hartley, J. and Sleeman, D. H. (1973) Towards intelligent teaching systems. *International Journal of Man-Machine Studies*, **5**: 215–236.

Heift, T. (1998) Designed Intelligence: A Language Teacher Model. Ph.D. thesis, Simon Fraser University.

Heift, T. (2001) Intelligent Language Tutoring Systems for Grammar Practice. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, **6**(2): 1–15.

Heift, T. (2003) Multiple learner errors and meaningful feedback: A challenge for ICALL systems. *CALICO Journal*, **20**(3): 533–548.

Heift, T. (2004) Inspectable Learner Reports for Web-based Language Learning. *ReCALL*, **16**(2): 416–431.

Heift, T. (2005) Corrective Feedback and Learner Uptake in CALL. *ReCALL*, **17**(1): 32–46.

Heift, T. and Schulze, M. (2003) Error diagnosis and error correction in CALL. *CALICO Journal*, **20**(3): 433–436.

Heift, T. and Schulze, M. (2007) *Errors and Intelligence in Computer-Assisted Language Learning: Parsers and Pedagogues*. New York: Routledge.

Hémard, D. (2004) Enhancing Online CALL Design: the case for evaluation. *ReCALL*, **16**(2): 502–519.

Hémard, D. (2006) Evaluating Hypermedia structures as a means of improving language learning strategies and motivation. *ReCALL*, **18**(1): 24–44.

Holland, V. M., Kaplan, J. D. and Sams, M. R. (eds.) (1995) *Intelligent language tutors: theory shaping technology*. Mahwah, NJ: Lawrence Erlbaum Associates.

Karlsson, F., Voutilainen, A., Heikkilä, J. and Anttila, A. (eds.) (1995) *Constraint Grammar: A Language-Independent System for Parsing Unrestricted Text*. Berlin and New York: Mouton de Gruyter.

Klein, P. (1998) CALICO Journal software evaluation review of ‘‘El Corrector’’ (‘un medio fácil para corregir rápidamente el español’). *CALICO Journal*, **16**(1): 64–75.

Krashen, S. D. (1987) *Principles and Practice in Second Language Acquisition*. New York: Prentice Hall International.

Krashen, S. D. (1988) *Second Language Acquisition and Second Language Learning*. New York: Prentice Hall International.

Kuenning, G. (2005) International Ispell Web, Version 3.3.02. http://ficus-www.cs.ucla.edu/geoff/ispell.html.

Leaver, B. L. and Willis, J. R. (2004) *Task-Based Instruction in Foreign Language Education*. Washington, DC: Georgetown University Press.

Lee, J. and VanPatten, B. (1995) *Making Communicative Language Teaching Happen*. New York: McGraw-Hill.

Levy, M. (1997) *Computer-Assisted Language Learning: Context and Conceptualization*. New York: Oxford University Press.

L'Haire, S. (2004) Vers un feedback plus intelligent. Les enseignements du project Freetext. In: *Proceedings of the Journée d'étude de l'ATALA on NLP and Language Learning*. 1–12. http://www.u-grenoble3.fr/lidilem/talal/actes/JourneeTALAL-041022-lhaire.pdf.

L'Haire, S. and Faltin, A. V. (2003) Error diagnosis in the FreeText project. *CALICO Journal*, **20**(3): 481–495.

Lightbown, P. M. and Spada, N. (1999) *How languages are learned*. Oxford: Oxford University Press.

Long, M. H. (1988) Instructed interlanguage development. In: Beebe, L. (ed.), *Issues in Second Language Acquisition: Multiple perspectives*. New York: Newbury House, 115–141.

Long, M. H. (1991) Focus on form: A design feature in language teaching methodology. In: Bot, K. D., Kramsch, C. and Ginsberg, R. (eds.), *Foreign language research in cross-cultural perspective*. Amsterdam: John Benjamins, 39–52.

Long, M. H. (1996) The role of linguistic environment in second language acquisition. In: Ritchie, W. C. and Bhatia, T. K. (eds.), *Handbook of second language acquisition*. New York: Academic Press, 413–468.

Lyster, R. (1998) Negotiation of form, recasts, and explicit correction in relation to error types and learner repair in immersion classroom. *Language Learning*, **48**: 183–218.

Martins, R., Nunes, G. and Hasegawa, R. (2003) Curupira: A Functional Parser for Brazilian Portuguese. In: *Computational Processing of the Portuguese Language. 6th International Workshop, PROPOR 2003, Faro, Portugal, June 26–27, 2003. LNCS 2721*. Berlin, Heidelberg: Springer Verlag, 179–183.

Michaud, L. N., McCoy, K. F. and Pennington, C. A. (2000) An Intelligent Tutoring System for Deaf Learners of Written English. In: *Proceedings of the Fourth International ACM SIGCAPH Conference on Assistive Technologies (ASSETS 2000)*. Washington, DC.

Michaud, L. N., McCoy, K. F. and Stark, L. A. (2001) Modeling the Acquisition of English: An Intelligent CALL Approach. In: *Proceedings of The 8th International Conference on User Modeling*. Sonthofen, Germany, 14–25.

Nagata, N. (1993) Intelligent Computer Feedback for Second Language Instruction. *The Modern Language Journal*, **77**(3): 330–339.

Nagata, N. (1996) Computer vs Workbook Instruction in Second Language Acquisition. *CALICO Journal*, **14**(1): 53–75.

Nagata, N. (1997) An Experimental Comparison of Deductive and Inductive Feedback Generated by a Simple Parser. *System*, **25**(4): 515–534.

Nagata, N. (2002) BANZAI: An Application of Natural Language Processing to Web based Language Learning. *CALICO Journal*, **19**(3): 583–599.

Nagata, N. (2009) Robo-Sensei's NLP-Based Error Detection and Feedback Generation. *CALICO Journal*, **26**(3): 562–579.

Norris, J. and Ortega, L. (2000) Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-Analysis. *Language Learning*, **50**(3): 417–528.

Ramsay, A. and Mirzaiean, V. (2005) Content-based support for Persian learners of English. *ReCALL*, **17**(1): 139–154.

Reuer, V. (2003) Error recognition and feedback with Lexical Functional Grammar. *CALICO Journal*, **20**(3): 497–512.

Schmidt, R. (1995) Consciousness and foreign language: A tutorial on the role of attention and awareness in learning. In: Schmidt, R. (ed.), *Attention and awareness in foreign language learning*. Honolulu: University of Hawaii Press, 1–63.

Schwind, C. B. (1995) Error Analysis and Explanation in Knowledge Based Language Tutoring. *Computer Assisted Language Learning*, **8**(4): 295–324.

Sleeman, D. (1982) Inferring (mal) rules from pupil's protocols. In: *Proceedings of ECAI 82*. Orsay, France, 160–164.

Swartz, M. L. and Yazdani, M. (eds.) (1992) *Intelligent tutoring systems for foreign language learning: The bridge to international communication*. Berlin: Springer.

Tschichold, C. (1999) Grammar Checking for CALL: Strategies for Improving Foreign Language Grammars. In: Cameron, K. (ed.), *CALL: Media, Design, and Applications*. Lisse: Swets and Zeitlinger, 203–222.

Tschichold, C. (2003) Lexically driven error detection and correction. *CALICO Journal*, **20**(3): 549–559.

Vandeventer, A. (2001) Creating a grammar checker for call by constraint relaxation: a feasibility study. *ReCALL*, **13**(1): 110–120.

Weischedel, R. M., Voge, W. M. and James, M. (1978) An Artificial Intelligence Approach to Language Instruction. *Artificial Intelligence*, **10**(3): 225–240.

Wiggins, G. (1998) *Educative Assessment*. San Francisco, CA: Jossey-Bass Inc.

Willis, D. and Willis, J. (2007) *Doing Task-based Teaching*. Oxford: Oxford University Press.