Computer Assisted Language Learning
Vol. 21, No. 4, October 2008, 323–338

From recording linguistic competence to supporting inferences about language acquisition in context

Luiz Amaral[a]* and Detmar Meurers[b]

[a]*Department of Languages Literatures and Cultures, University of Massachusetts, Amherst, USA;* [b]*Department of Linguistics, Universität Tübingen, Germany*

(*Received 18 February 2008; final version received 3 June 2008*)

Student models for Intelligent Computer Assisted Language Learning (ICALL) have largely focused on the acquisition of grammatical structures. In this paper, we motivate a broader perspective of student models for ICALL that incorporates insights from current research on second language acquisition and language testing. We argue for a student model that includes a representation of the learner's ability to use language in context and to perform tasks, as well as for an explicit activity model that provides information on the language tasks and the inferences for the student model they support. The student model architecture we present is being developed as part of the TAGARELA system, an intelligent workbook supporting the instruction of Portuguese.

Keywords: Intelligent Computer-Assisted Language Learning (ICALL); intelligent tutoring systems (ITS); natural language processing (NLP); second language acquisition (SLA); student model; learner model; activity model; task appropriateness; task strategies; L1 transfer

Introduction

Research in Intelligent Computer Assisted Language Learning (ICALL), the subfield of CALL integrating natural language processing and artificial intelligence, has typically equated language acquisition with learning of linguistic patterns and grammatical structures. ICALL systems therefore have focused on identifying and providing feedback to problems in student performance that are exclusively related to linguistic structures. Some of these systems include learner models to adapt explanations, offer advice or decide on how a given student can advance through the material. But the view of second language acquisition focusing on absolute linguistic abilities has hindered the development of learner models that take into consideration the ability of students to perform language tasks; the strategies they must master to successfully use language in context; and their linguistic abilities relative to the linguistic context and the task.

The purpose of this paper is to advance the conceptualization of ICALL student models, primarily by integrating a more comprehensive view of the language acquisition process. We start by reviewing student models used in ICALL in recent years. Then we discuss the basis on which we want to advance the conceptualization of student models for ICALL, a model of language acquisition that is in line with current SLA research and an

*Corresponding author. Email: amaral@spanport.umass.edu

ISSN 0958-8221 print/ISSN 1744-3210 online

DOI: 10.1080/09588220802343454
http://www.informaworld.com

understanding of how evidence can support inferences about student competence that is informed by language testing research. To make these ideas concrete, in section four we introduce the ICALL system TAGARELA, a web-based ICALL system accompanying the instruction of Portuguese. On this basis, in section five we discuss how the extended conceptualization of the student model enhances the way such a system can provide feedback to the learner.

Current student models in ICALL

ICALL systems that have been incorporated into foreign language teaching practice present either a variety of activities for different language topics (cf., e.g., Heift, 2003; Nagata, 2002; Hagen, 1999; Murphy & McTear, 1997), or they support students in editing and correcting their writing (cf., e.g., Michaud, McCoy, & Pennington, 2000). In this section, we briefly review the learner models of three representative systems, paying particular attention to the assumptions about the language acquisition process underlying these learner models.

E-Tutor

The E-Tutor is an ICALL system that functions as an electronic workbook for German students at Simon Fraser University. E-Tutor is fully integrated into the language program and several papers have been written about its development and use (cf., e.g., Heift, 2003, 2004; and the article in this issue). The student model of the E-Tutor keeps track of individual ‘grammar skills’ a student is acquiring, such as subject-verb agreement or the subcategorization and form of arguments. The system collects positive and negative evidence from the student’s production and uses it to keep an absolute score of the student’s knowledge of each individual grammar skill. There is no overall classification of a student’s grammatical competence; the learner model keeps track of the student’s level of performance per grammar skill. To classify a student’s knowledge, the numeric performance scores are grouped into three levels: beginner, intermediate, and advanced. When the system identifies a specific grammatical error in the student’s input, it checks the level of proficiency of that student for the particular grammar skill concerned. The system then decides which feedback message to use.

E-Tutor’s learner model is not based on a specific theory of SLA. It models the learner’s language acquisition as the acquisition of individual ‘grammar skills’, which are modeled independently, without interconnection among them, using absolute scores for each one.

ICICLE

ICICLE (Michaud et al., 2000) is a system developed to provide writing assistance to native speakers of American Sign Language learning English as a second language. It receives short English essays written by the students and provides tutorial feedback on grammatical errors. The goal of the student model (Steps of Language Acquisition in a Layered Organization Model (SLALOM)) (Michaud, McCoy & Stark, 2001) is to capture the status of the grammatical structures of English in terms of ‘acquired’, ‘being-acquired’, and ‘unacquired’. The knowledge units of the SLALOM architecture are grammatical concepts connected to two sets of grammar rules, English rules and mal-rules. These knowledge units are grouped and hierarchically classified. The hierarchy proposed is based on studies in second language acquisition that showed stereotypical sequences in the acquisition of grammatical structures by learners of the same L1 community (cf., e.g., Gass, 1979; Krashen, 1988; Schwartz & Sprouse, 1996). This hierarchy is used to identify the current state of knowledge of a learner and to predict the next grammatical structures to be acquired. The system compares stereotypical sequences of English acquisition with the student's current production to create an analysis of what structures are being acquired.

Different from the E-Tutor, SLALOM is based on explicit claims about how grammatical structures are acquired by the learner. At the same time, the SLA process being modeled is restricted to the acquisition of linguistic properties. Since ICICLE's goal is to serve as a writer's aid this may be sufficient given that interaction with the learner is limited to providing feedback to a specific single type of activity, essay writing.

CASTLE

CASTLE (Murphy & McTear, 1997) provides communicative role-play scenarios where students answer questions posed by the system. When the system diagnoses certain linguistic problems with a student's production (i.e., when a student makes three errors of the same type), it proposes a set of remedial exercises. CASTLE's learner model stores information in three distinct groups. The first one is the student's personal information, which includes native language, motivation, background language proficiency, etc. The second is the 'student model', which keeps track of the student's performance by 'domain topics', her proneness to commit certain errors, and the likely causes of errors. The third is the 'cognitive model', which stores information about the student's preferred feedback media and exercise types, her interest in grammar, and the use of polite forms. Note that the 'student model' of CASTLE is the part of its learner model that keeps track of the student's proficiency level in terms of her linguistic performance.

CASTLE also uses stereotypical sequences of language acquisition to update its student model. Different from SLALOM's use of stereotypical sequences from SLA research, CASTLE orders the acquisition of grammar in terms of 'grammatical partitions' based on typical textbook progressions in foreign language teaching.

Conceptually speaking, CASTLE's learner model differs from the two models presented above because it acknowledges the needs of ICALL systems to have information about a student that is not restricted to her grammatical performance. By keeping track of a student's preferences and background, the system is able to take some characteristics of the learner into account when providing feedback. However, the additional information modeled in CASTLE is completely dissociated from the student's domain knowledge. The information is not referenced when interpreting or recording errors which the system identifies in the student input.

Advancing student models for ICALL

Modeling language acquisition

To fulfill their functions within a given ICALL system, student models have to represent the relevant information about the learner's acquisition process. This process has been extensively explored and characterized in second language acquisition research, of which we highlight some of the relevant key aspects here. While we do not assume the presence of a specific linear progression in the acquisition of an L2, language acquisition clearly is a process. The development of the interlanguage is gradual, and some structures are acquired before others. Learner models need to capture as precisely as possible this process so that we can arrive at an accurate description of a learner's abilities.

Characterizing L2 acquisition in general, Ellis (2003) concisely stated that 'the general goal of language learning is the fluent, accurate, and pragmatically effective use of the target language'. According to Canale and Swain (1980), the four major types of competence a learner needs to acquire to achieve this goal are grammatical competence, sociolinguistic competence, discourse competence and strategic competence. Starting in the 1980s, the role of learning strategies in the SLA process was reviewed and emphasized (cf., e.g., Canale, 1983; O'Malley & Chamot, 1990). Bachman (1990) reviews the literature on communication strategies, and re-conceptualizes strategic competence as a basic cognitive management function for language use. Essentially, strategic competence can be understood as the set of non-linguistic properties used by the learner that play a role in language production.

One example of strategic competence cited by Bachman and Palmer (1996) is planning. According to them, when a student is planning how to approach a test task, she has to:

(i) select 'elements from the areas of topical knowledge and language knowledge for successfully completing the task';
(ii) formulate 'one or more plans for implementing these elements in a response to the test task'; and
(iii) select 'one plan for initial implementation as a response to the test task'.

Different from Bachman and Palmer (1996), we are not specifically interested in strategies used by students when taking a test. We want to model the ability of students to handle specific language tasks in order to identify possible sources of interference in the student's production that are not necessarily related to her explicit awareness of the target language forms and rules. For example, situations where a student commits a syntactically motivated error, such as agreement or subcategorization errors, while doing an exercise which requires the use of a specific strategy to complete a meaning-based task, such as a reading comprehension question where the student has to scan the text for specific information. In such a case, the error may be motivated by the student's lack of knowledge of the syntactic properties in question, or it may be influenced by the complexity of the linguistic context or the task.

Complementing this research into the non-linguistic abilities that need to be acquired by the learner, researchers in language testing have also paid close attention to the role of the environment of the student's production. Considering the nature of the assessment of learner performance, Mislevy (2006) highlights that the 'interpretation of [a student's] actions rather than the actions themselves constitute data in an assessment argument'. He continues by presenting the three kinds of data that provide information necessary to interpret a student's production and create the 'assessment argument':

'(i) aspects of the situation in which the person is acting;
(ii) aspects of the person's action in the situation; and
(iii) additional information about the person's history or relationship to the observational situation.'

The influence of the situation on a student's production in an ICALL system is directly related to the properties of the tasks that constitute the exercises presented by the system. The notion of a task has been defined in various ways in the SLA literature and it is instructive to review some of these definitions as basis of a characterization appropriate for ICALL system development. Ellis (2003, pp. 4–5) presents a list of some common definitions of task. Richards, Plat and Weber (1985), for example, describe task as

> an activity or action which is carried out as the result of processing or understanding language, i.e., as a response. For example, drawing a map while listening to a tape, and listening to an instruction and performing a command may be referred to as tasks. Tasks may or may not involve the production of language.

Skehan (1996) defines a task as being 'an activity in which: meaning is primary; there is some sort of relationship to the real world; task completion has some priority; and assessment of task performance is in terms of task outcome'. For Nunan (1989) a task is

> a piece of classroom work which involves learners in comprehending, manipulating, producing, or interacting in the target language while their attention is principally focused on meaning rather than on form. The task should also have a sense of completeness, being able to stand alone as a communicative act in its own right.

Bygate, Skehan and Swain (2001) argue that 'a task is an activity which requires learners to use language, with emphasis on meaning, to attain an objective'. Integrating Bygate et al. (2001)' s perspective with the role of instructions and the relevance of a connection to the real world mentioned above, for our ICALL research we will define a task as a contextualized activity that requires the learner to process both linguistic and non-linguistic information in a meaningful way following a specific set of instructions in order to produce language output. In the case of the TAGARELA system, which will be introduced in more detail later, the tasks that involve reading texts and listening to audio passages are primarily meaning-based; a meaningful connection to real-world situations is also incorporated in more constrained activities which focus on language forms.

In the design of the student model, we integrate these three strands: our definition of tasks for ICALL systems; Bachman's conceptualization of strategic competence; and Mislevy's considerations about the role of the environment in interpreting a student's production. On this basis, we can design a model of learner competence that can provide enough information for a system to make inferences about the student's grammatical knowledge as well as his ability to use this knowledge to complete different tasks in different environments. We propose that such model be a combination of two parts:

(i) linguistic competence as a model of the acquisition of linguistic structures relative to the environment of language use; and
(ii) strategic competence as a model of the acquisition of non-linguistic properties relevant to language use.

As we will show in our discussion of the benefits of an extended learner model later, taking into consideration the activity environment to draw inferences about a student's linguistic competence allows the system to react more appropriately to certain types of learner errors.

Assessing learner abilities to build a student model

Student models are built and modified based on observations of learner performance (or using information explicitly provided by the learner). In line with Mislevy's comment, this requires interpretation of the student production. Student models thus store information inferred about the abilities the student used to produce a given sentence.

Research in ICALL has paid little attention to the validity of the inferences made by the system about a student's current state of abilities. It is usually taken for granted that linguistic errors are caused solely by a lack of linguistic competence. In contrast, we want to shift the perspective to take into account the fact that the task being performed can play a significant role in determining the students' production. To build a model that takes into account the linguistic and the strategic competence of a student, it is necessary to provide mechanisms ensuring that the system's inferences about a student's state of abilities are valid. To further motivate and ground our perspective, let us take a look at some related research on validity in language testing.

Describing the concept of validity for language tests, Bachman and Palmer (1996) state that 'construct validity pertains to the meaningfulness and appropriateness of the interpretations that we make on the bases of test scores' and that 'in order to justify a particular score interpretation, we need to provide evidence that the test score reflects the area(s) of language ability we want to measure' (Bachman & Palmer, 1996, p. 21). Assigning and interpreting test scores is a similar process to describing a student's current state of abilities. In fact, scoring is a mechanism commonly used in student models to identify levels of proficiency for specific knowledge units. The issue of the validity of test scores thus applies directly to the validity of the information in a student model.

In the case of ICALL systems that present specific exercises, there are two issues related to the validity of system inferences that we need to pay particular attention to. The first one is known as content validity, which McNamara (2000, p. 50) characterizes as the concept that explains the 'extent to which the test content forms a satisfactory basis for the inferences to be made from test performance'. For ICALL system design, this means that it is important to ensure that the exercise types and contents offered by the system are sufficient to make the necessary inferences about students' competence.

The second issue on validity of inferences to be highlighted here relates to the methods used to obtain information about students' linguistic competence. There are two ways in which properties of exercises affect the result of the system's observations, which we can also characterize using notions from assessment theory (cf., e.g., McNamara, 2000). Construct irrelevant variance occurs when a given exercise introduces factors that are not relevant to measure the ability we want to observe. Construct under-representation occurs when the exercise is too easy for the student, jeopardizing the observation of a given ability. Particular care needs to be taken when the skill observed is embedded in contexts that are unfamiliar to the student or irrelevant to what is being assessed. Bachman and Palmer (1996, p. 21) emphasize that the analysis of a student's performance has to be interpreted with respect to a 'specific domain of generalization'. Thus, when we consider the validity of an interpretation, 'we need to consider both the construct definition, and the characteristics of the test tasks'.

In sum, in order to guarantee valid interpretations of student performance it is not enough to keep track of a student's production; it is vital to have at least information about the environment where it occurs. Without a clear description of the exercise items that triggered the student performance, our interpretations about levels of proficiency may not be accurate.

Our approach

The TAGARELA system

TAGARELA (Teaching Aid for Grammatical Awareness, Recognition and Enhancement of Linguistic Abilities) is a web-based ICALL system accompanying the instruction of

Portuguese. The system was designed to help deal with some of the limitations of the foreign language classroom environment, such as the limited time instructors have to provide on the spot individualized feedback. The system can be viewed as an intelligent electronic workbook. Its activity types are similar to the ones found in traditional workbooks, and are divided into six groups: reading, listening, description, rephrasing, fill in the blanks, and vocabulary. The expected input consists of words, phrases or sentences. Different from traditional workbooks, TAGARELA offers on-the-spot individualized feedback on spelling, morphological, syntactic and semantic errors. Answers to all activities are electronically checked, i.e., the generation of feedback is completely automated. TAGARELA can be used as a pedagogical complement in traditional classroom settings, as well as in distance learning or individualized instruction programs.

The general TAGARELA architecture shown in Figure 1 consists of six modules: Interface, Analysis Manager, Feedback Manager, Expert Module, Instruction Model and Student Model. The Analysis Manager receives the input sentence and gathers the necessary information from the Instruction and Student Models to decide on the best processing strategy (i.e., which submodules to call in which order). It then calls the appropriate sub-modules in the Expert Module to analyze the input. The tokenizer takes into account specifics of Portuguese, such as cliticization, contractions and abbreviations.

The full-form lexical lookup then returns multiple analyses based on the CURUPIRA lexicon (Martins, Hasegawa, & Graças Nunes, 2006), including detailed morphological information. In the spirit of Constraint Grammar (Karlsson, Voutilainen, Heikkilä, & Anttila, 1995; Bick, 2000, 2004), finite state disambiguation rules are used to narrow down

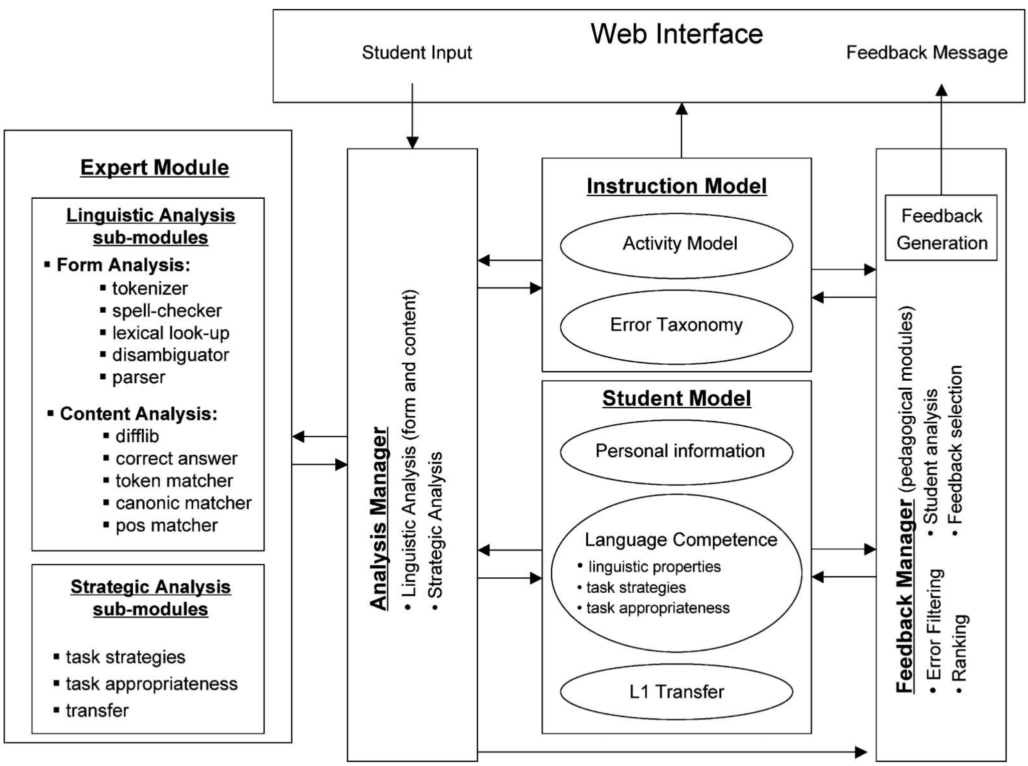

Figure 1. Architecture of the TAGARELA system.

the multiple lexical analyses based on the local context. Complementing these local disambiguation rules, a simple bottom-up chart parser using hand-written rules checks agreement and case relations and some global well-formedness conditions. In addition to the form-focused processing, content assessment is performed using shallow semantic matching between the student answer and teacher targets from the Activity Model, essentially a basic version of the approach discussed in Bailey and Meurers (2008). The result of processing is a representation of the learner input that is annotated with the properties obtained through the natural language processing (NLP) analyses. The annotated input is then passed on to the Feedback Manager. The Feedback Manager receives the annotated input, gathers information about the student and the activity, filters the errors that should be targeted and decides on the best feedback message to generate. The explicit Instruction Model and the Student Model are the repository of information that complements the information obtained by the NLP of the learner input and guides the processing mechanism from linguistic analysis to feedback generation.

Aspects of the activity model of TAGARELA

We established earlier that 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 task in which 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 task.

In TAGARELA, each activity includes an explicit activity model specifying two types of information. First, all activities are classified based on three criteria: level; nature of input; and content manipulation. The first one is the level of the activity based on its course number and the module/lesson number. We take the progression of the course material to be an indicator of the complexity of the linguistic forms necessary to perform specific tasks.[1] The second criterion, nature of input, reflects the three types of input accepted by the system: (i) word, (ii) phrase, (iii) sentence. The nature of the input criterion is used by the system to identify the complexity of possible answers. The more linguistically complex the required target structure, the more difficult it is for a student to complete the task. Finally, the third criterion encodes the amount of content manipulation necessary to fulfill the task. There are four categories that encompass the six activity types: (i) little content manipulation; (ii) some content manipulation; (iii) necessary content manipulation; and (iv) major content manipulation. In general, the amount of content manipulation could be linked to a specific activity type, e.g., fill-in-the-blanks typically require little content manipulation, while most reading and listening exercises generally require major manipulation of content to be appropriately answered. However, individual questions may present significant differences within the same activity type, so having a separate criterion for content manipulation allows us to explicitly register these differences and better relate the occurrence of form-based errors to the meaning-based requirements of the exercise where the errors occur. Content manipulation is also an important measure because it is a main distractor from a focus on grammatical accuracy. This classification implies that the greater the amount of expected content manipulation by the student in a task, the more likely she will be to make grammatical errors.

The second type of information encoded in the activity model is about possible inferences the system can make based on each individual question. For example, while a rephrasing question may provide enough information for the system to observe if the student can deduce the necessary syntactic rules to re-write the target sentence, a 'wh-'

question in a reading comprehension activity may allow the system to observe the inferences a student needs to make to correctly answer the question. The activity model explicitly provides the system with information about how to interpret a student's production based on the type of task the student is performing. Note that this information cannot be derived from the three criteria discussed above for classifying activities because different questions for the same activity type may or may not allow the system to draw inferences about a given knowledge unit.

The *strategic analysis submodules* use the information provided by the activities to diagnose sources of errors in the student input. The *task appropriateness submodule* uses the description of activities presented above to classify linguistic errors in relation to the type of activity where they occur. The *task strategies submodule* draws inferences about the possible causes of an error based on information about each individual question, as described in the previous paragraph. Finally, the *transfer submodule* uses a list of false cognates (English–Portuguese) to compare individual target tokens with unmatched tokens in the student input to check if errors classified as 'missing content word' can be caused by negative lexical transfer. Notice that, with the exception of the transfer submodule, the other submodules of the strategic analysis use a combination of linguistic information about the input provided by the linguistic analysis submodules and information about the activity provided by the activity model. In the next section, we will see where the analysis provided by these submodules is stored in the student model.

Taken together, the classification of activities in the Instruction Model in combination with the analysis of the input produced by different expert modules and annotated by the Analysis Manager allows the system to model important aspects of the student's linguistic and non-linguistic competence involved in language acquisition.

The student model of TAGARELA

The proposed student model consists of three repositories: Personal Information; Language Competence; and L1 Transfer. We here focus on describing what information is stored where and turn to discussing the benefits of the extended student model in the section afterwards.

- *Personal Information*: The first repository stores the student's personal information, such as name, level, age, gender and native language. Some of this information is used by the system when providing feedback, but most of it is only stored for potential future use, such as second language acquisition research that might require this type of information.
- *Language Competence*: The second repository stores information about the student's language competence. As motivated earlier, we subscribe to a broad view of language competence that encompasses not only linguistic competence, i.e., knowledge about language forms and rules, but also the relevant non-linguistic abilities that have to be developed by the learner to use language in order to perform the tasks in TAGARELA. In other words, the language competence model keeps track of linguistic properties observed that are relative to the task performed and the intra- and extra-linguistic context. Concretely, the system stores three types of information under Language Competence: Linguistic Properties, Task Appropriateness, and Task Strategies.
- *Linguistic Properties:* The Linguistic Properties are divided into *form-driven* and *content-driven*, reflecting two types of linguistic analysis that are performed by the

system's NLP modules: form analysis of spelling and morpho-syntactic errors, and shallow content analysis providing information about the semantic appropriateness of the input. The form-driven properties that can be observed and identified by the NLP include spelling, determiner-noun and subject-verb agreement, and word order properties determined by the syntactic processing. Content-driven properties can be the result of extra/missing content words, wrong selection, word choice or collocation, negative lexical transfer, required concept matching or synonym identification.

- *Task Appropriateness*: Task Appropriateness stores information about the performance of the student relative to the activity classification. As presented in the previous subsection, each activity is classified in terms of its type (e.g., listening, reading), level, nature of the input (word, phrase, sentence) and complexity of content manipulation required. The system uses this classification to store information about the student's performance relative to these distinctions.
- *Task Strategies:* The third type of information under Language Competence is related to a student's performance in relation to task strategies. This part of the student model keeps track of specific strategies students have to use to complete a given activity, e.g., scanning a text to locate specific information or getting the gist of listening passages. As noted in the previous subsection, information about necessary strategies to complete a given task is hand-specified in the activity model.
- *L1 Transfer:* The last repository of the student model is dedicated to L1 transfer errors. The L1 Transfer component stores information about indicators of possible structural and lexical transfer errors from the native language of the student into the second language (cf., e.g., Odlin, 2003). The diagnosis of possible transfer errors is performed by the specific processing submodule of the Expert Module, as included in Figure 1.

Using the information from the student model

As mentioned in the introduction, the TAGARELA system provides individual feedback based on the students' input to an exercise. Feedback is provided on the semantic appropriateness as well as on the grammatical accuracy of the input. The choice of the feedback strategy and contents is based on the student input, the activity model for the exercise the student is dealing with, and the student model. The general feedback strategy uses metalinguistic messages and scaffolding techniques to help the learner develop self-editing skills (cf., e.g., Hyland & Hyland, 2006).

Most relevant here is how the content of the feedback message is determined. The content depends on identifying the likely source of the error. Based on the learner input annotated by the NLP modules, the student model, and the activity model, the system distinguishes between three possible error sources.

Firstly, an error can result from a student's lack of a specific linguistic ability, e.g., when a given student has not mastered subject-verb agreement. This is the classic case handled by ICALL systems, whereas the next two rely on the extended student model proposed in the previous subsection.

Secondly, an error can result from the student's lack of a strategic ability needed for a given task. For example, if the learner has problems scanning a text to locate the relevant concepts, they cannot correctly answer a reading comprehension question asking for those concepts. To diagnose such an error, the system compares the concepts that the activity model identifies for a given text with the corresponding concepts identified in the learner input by the NLP modules. The learner model provides the information whether the learner has been able to pick up the relevant concepts in reading comprehension before.

Thirdly, an error can result from an insufficient mastery of a specific linguistic ability, which allows the student to use it only in certain tasks or constructions. For example, a student may be able to formulate simple sentences with correct subject-verb agreement as part of a picture description task, but fails to use correct agreement forms when answering listening comprehension questions that require more complex content, a complex form or an otherwise increased cognitive load. As in the previous case, the student model and the activity model are essential for determining whether the problem lies in the use of the linguistic forms in general or whether there is a correlation with the use of these linguistic forms only in particular tasks.

Exemplifying the benefits of an extended learner model

Having motivated a conceptualization of student models for ICALL that includes the task performed by the student and the strategies required to complete the task, and the concrete intelligent workbook TAGARELA on the basis of which our work is based, we can now turn to making explicit where the extended student models have a concrete impact on the ability of an ICALL system to react to student input.

Task strategies

We first discuss the relevance of including task strategies in the student model. Consider a student answering reading comprehension questions. By processing the student input to a number of these questions and comparing them to the target representations in the activity model, the system can determine that the student repeatedly failed to include key concepts in the answers. Based on the activity model, the system determines that these questions required scanning the text for specific information. As a result of having identified the recurring difficulty of including all the concepts in activities requiring scanning of a text for the information, the next time a concept is missing in the learner input for an activity with similar properties, the system can provide a feedback message that targets the student's problem with scanning. Such a message could be of the following kind: 'Try to scan the text more carefully to include all the key concepts in your answer.'

To see where such feedback differs from that given by an ICALL system that focuses exclusively on linguistic abilities, consider the options for such a system. Given the input, it can determine that the student has problems including all the words in the answer that are specified in the target answers. Using lexical resources or a part-of-speech tagger to analyze the learner input and the targets, the system can also determine that the word missing from the student input is a noun. By keeping track of this information in a student model storing the global linguistic abilities of the student, the system can determine when a given student repeatedly leaves out required nouns from the answer. On this basis, the system can then provide as feedback something such as: 'There is a noun missing in your sentence again. Please make sure to include all of them in your answer.' Such a student model recording only the global linguistic abilities of the student does not support reference to the type of activity and the strategies required by that activity which the student has repeatedly had problems with.

Task appropriateness

For the second issue, task appropriateness, it is useful to keep in mind that it is independent of the relevance and role of the task strategies discussed in the previous section, i.e., the extra-linguistic strategies that students need to master to provide correct answers. Task appropriateness instead encodes whether a specific linguistic ability has been observed in a given task, i.e., it makes the recording of linguistic ability relative to the environment in which that ability has been observed.

Consider a situation where a specific student has provided input to several activities. By analyzing the student input, the system can determine that in most fill-in-the-blank activities where the student needs to enter the correct finite verb form, the student is able to produce finite verb forms that correctly agree with the subject provided in the activity. The same student has also provided some full sentence answers to reading comprehension activities. Analyzing those answers, the system determines that they frequently do not show correct subject-verb agreement. Based on the student model obtained from those interactions, the system can infer that the student is in principle aware of the need to encode subject-verb agreement, but has problems doing so when the entire sentence needs to be created (or that the student's attention might be consumed by the content-manipulation needs of an activity or other cognitive demands). What exactly the system can infer depends on what it can determine about the activity, i.e., generally it depends on the explicitness of the activity model).

As a result of being able to record linguistic performance relative to the task and linguistic environment, the extended student model can change the prioritization of feedback provided by the system and it can change the nature of the feedback. In our example, the system can give different priority to reporting subject-verb agreement errors depending on the activity type or level.

On the other hand, the standard ICALL system with a student model including only an absolute record of the linguistic performance can only determine that the student sometimes has problems with subject-verb agreement. It does not have access to the information needed to prioritize or phrase feedback differently, relative to the task type, level, linguistic environments or other properties made explicit in the activity model.

L1 transfer

Whereas the previous two aspects extend the learner model with a representation of the task performed by the learner and allow the system to interpret student input on the basis of the extralinguistic and linguistic requirements for completing that task, the third learner model extension we proposed focuses on extending the set of individual properties of the learner that are represented; that is, individual properties of the learner which can support additional inferences by the system when interpreting student input.

We focus on the native language, the L1 of the student, as an important property to be reflected in the student model. However, any measures of the student's language learning or, more generally, any measure of the student's cognitive abilities with predictive power on language learning (cf., e.g., Chun & Payne, 2004) should be considered for inclusion in the student model.

Here we exemplify the relevance of storing the L1 in the student model by considering a Portuguese native speaker as student of English.[2] Based on the information about the L1 and a representation of typical lexical transfer errors, i.e., false cognates, for this combination of L1 and L2, it is possible to support stronger inferences based on the student input.

For example, in answering a comprehension question, a Portuguese learner of English writes: 'John assumed Bill was wrong' for a question where the target response is 'John admitted Bill was wrong'. Given that in Portuguese the word 'assumir' is translated as

English 'admit' and not the false friend 'assume', the system can determine that the problem is likely to be negative lexical transfer and base the feedback on this specific diagnosis.

A baseline system which cannot make reference to a student model encoding the L1 of the student as well as a representation of negative lexical transfer opportunities given this L1–L2 pair will not be able to infer why the student used a word expressing the wrong meaning. Did the student not understand the text? Did she not understand the question? Did she make a wrong lexical choice in formulating the answer? The extended student model is needed to disambiguate these possibilities and draw the inference that this is likely to be an instance of L1 transfer, which makes it possible to prioritize feedback on lexical transfer errors over a general meaning error.

Related work

At the beginning, we discussed the student models of three representative ICALL systems and argued that limiting the model to the acquisition of linguistic competence can be problematic for an ICALL system. The general impetus of our argument is closely related to that of Bull, Brna and Pain (1995), who argue for extending the scope of student models to incorporate aspects outside the boundary of the linguistic domain knowledge. While the two proposals share the general direction, they differ in the nature and the interaction of these extensions with the domain knowledge. Bull et al. (1995) propose adding information about learning strategies and analogy in two modules that are separate from their model of the acquisition of linguistic properties. This separation is intended to emphasize the general nature of analogy and learning strategies across different domains. Bull et al. (1995) acknowledge the fact that different domains require different types of analogies and learning strategies, but their architecture posits three separate modules and does not address the question of linking specific aspects of the domain knowledge (or the language tasks in which they are needed) to specific analogies or learning strategies. For expository reasons, this example uses the inverse scenario of TAGARELA, which is used by Americans learning Portuguese. This might also be a result of the fact that the sample application discussed targets a single linguistic phenomenon, clitic pronoun placement in European Portuguese, and the paper does not mention activity design. As a result, their proposal does not address how different language activities, requiring correlated subsets of linguistic competence and analogies, can be encoded.

For our proposal and the realization of it as part of the TAGARELA system, on the other hand, the use of language to perform different language tasks in a real-life learning environment is a central design element. As a result, the extended learner model we propose incorporates task strategies and task appropriateness as core components of the language acquisition process. This is possible because the TAGARELA architecture supports correlating the linguistic features used by the learner with information about where they are used and for which purposes. The Expert Module produces an analysis of the features to be acquired that can be correlated with information explicitly encoded in the Activity Model allowing the system to make specific inferences about the acquisition of language in a task-based perspective.

Conclusion

In this paper, we motivated the need to develop student models for ICALL which go beyond the acquisition of grammatical competence. We argued for extending ICALL student models beyond absolute linguistic knowledge, to include, firstly, the learner's abilities to use language in context, using appropriate strategies for specific goals; secondly, the learner's abilities relative to task type and complexity; and thirdly, the possibility of L1 transfer. The extended learner model reflects current theories of second language acquisition and allows the system to react to learner input as part of meaningful language tasks. We then exemplified the conceptual arguments on the basis of an intelligent web-based workbook, the TAGARELA system. The first version of the system has successfully been used with language learners in several courses at the Ohio State University. The extended version of the student model proposed in this paper, which was motivated in part by the feedback we received from students using the first version of the system, is currently under development.

While this paper has focused on motivating an extended conceptualization of student models for ICALL, we have also included a discussion of the concrete advantages of a system with access to such extended student models. Once the extended learner model is fully realized, we intend to follow up on this discussion of qualitative differences with experimental results comparing student groups using the TAGARELA system with different student models to empirically validate the impact of the extended student model in a real-life language learning environment.

Updating the student model currently requires significant hand-specification of the explicit activity models. Explicit activity models are well motivated by the need to support valid inferences about the student's state of knowledge. As we argued in Amaral and Meurers (2007b), explicit activity models are also well suited for a demand-driven NLP architecture supporting a wider range of activities. We intend to explore how to derive some of these activity properties automatically via additional NLP analyses.

Acknowledgements

For helpful discussion and suggestions, we would like to thank the ICALL research group (http://purl.org/net/icall) at the Ohio State University, where we developed the work reported here: Susan Bull, Trude Heift, Donna Long, Kathy McCoy, Bob Mislevy, Scott Payne, Mathias Schulze and the reviewers of UserModeling 2007, where an earlier version of our proposal was presented (Amaral & Meurers, 2007a).

Notes

1. This simplification was necessary at this stage of the project development for purely practical reasons.
2. For expository reasons, this example uses the inverse scenario of TAGARELA, which is used by Americans learning Portuguese.

Notes on contributors

Luiz Amaral (http://people.umass.edu/amaral/) is an assistant professor of Hispanic Linguistics at the University of Massachusetts Amherst. He designed with Detmar Meurers the TAGARELA system (http://tagarela.osu.edu), an ICALL system for the instruction of Portuguese. He works with activity design and natural language processing of learner language for intelligent language tutoring systems. He is also interested in student models, and lexical acquisition by second language learners.

Detmar Meurers (http://purl.org/dm) is a professor of computational linguistics at the University of Tübingen and an adjunct associate professor at the Ohio State University. In 2003, he started an ICALL research group (http://purl.org/net/icall) to explore the interface of natural language processing and real life foreign language teaching, with a focus on intelligent web-based workbooks supporting a range of meaning-based activities, learner modeling, and authentic text-based activity generation (http://purl.org/net/werti). To foster interdisciplinary collaboration as a crucial component of sustainable ICALL research, he co-organized workshops on the 'Interfaces of ICALL' (http://purl.org/net/iicall) and the 'Automatic Analysis of Learner Language' (http://purl.org/net/calico-workshop).

References

Amaral, L., & Meurers, D. (2007a). Conceptualizing student models for ICALL. In C. Conati, & K.F. McCoy (Eds.), *User Modeling 2007: Proceedings of the eleventh international conference* (pp. 340–344). Lecture Notes in Computer Science 4511. Wien, New York, Berlin: Springer. Retrieved September 1, 2008, from http://purl.org/dm/papers/amaral-meurers-um07.html

Amaral, L., & Meurers, D. (2007b). Putting activity models in the driver's seat: Towards a demand-driven NLP architecture for ICALL. Paper presented at the Symposium on NLP in CALL, EUROCALL 2007. University of Ulster, Coleraine Campus, Northern, Ireland. Retrieved September 1, 2008, from http://purl.org/net/icall/handouts/eurocall07-amaral-meurers.pdf

Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford, UK: Oxford University Press.

Bachman, L.F., & Palmer, A.S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford, UK: Oxford University Press.

Bailey, S., & Meurers, D. (2008). Diagnosing meaning errors in short answers to reading comprehension questions. In J. Tetreault, J. Burstein, & R.D. Felice (Eds.), Proceedings of the 3rd workshop on innovative use of NLP for building educational applications (pp. 107–115), held at ACL 2008. Columbus, Ohio: Association for Computational Linguistics. Retrieved September 1, 2008, from http://purl.org/dm/papers/bailey-meurers-08.html

Bick, E. (2000). *The parsing system "Palavras": Automatic grammatical analysis of Portuguese in a constraint grammar framework*. Aarhus, Denmark: Aarhus University Press.

Bick, E. (2004). PaNoLa: Integrating constraint grammar and CALL. In H. Holmboe (Ed.), *Nordic language technology, arbog for Nordisk sprogteknologisk forskningsprogram 2000–2004 (Yearbook 2003)* (pp. 183–190). Copenhagen: Museum Tusculanum.

Bull, S., Brna, P., & Pain, H. (1995). Extending the scope of the student model. *User Modeling and User-Adapted Interaction*, *5*, 45–65.

Bygate, M., Skehan, P., & Swain, M. (2001). *Researching pedagogic tasks, second language learning, teaching and testing*. Harlow: Longman.

Canale, M. (1983). On some dimensions of language proficiency. In J.W. Oller Jr. (Ed.), *Issues in language testing research* (pp. 333–342). Rowley, Mass: Newbury House.

Canale, M., & Swain, M. (1980). Theoretical basis of communicative approaches to second language testing. *Applied Linguistics*, *1*, 1–47.

Chun, D.M., & Payne, J.S. (2004). What makes students click: Working memory and look-up behavior. *System*, *32*(4), 481–503.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford, UK: Oxford University Press.

Gass, S. (1979). Language transfer and universal grammatical relations. *Language Learning*, *29*(2), 327–344.

Hagen, L.K. (1999). *Spanish for business professionals*. Project web page. Retrieved September 1, 2008, from http://www.uhd.edu/academic/research/sbp/

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 Journal*, *16*(2), 416–431.

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 1–46.

Karlsson, F., Voutilainen, A., Heikkilä, J. & Anttila, A. (Eds.). (1995). *Constraint grammar: A language-independent system for parsing unrestricted text*. Berlin and New York: Mouton de Gruyter.

Krashen, S.D. (1988). *Principles and practice in second language acquisition*. New York: Pergamon Press.

Martins, R., Hasegawa, R., & Graças Nunes, M.das. (2006). Curupira: a functional parser for Brazilian Portuguese. *Computational processing of the Portuguese language, 6th international workshop*. *Propor*. *lecture notes in computer science 2721*. Faro, Portugal: Springer.

McNamara, T. (2000). *Language testing*. Oxford, UK: Oxford University Press.

Michaud, L.N., McCoy, K.F., & Pennington, C.A. (2000). An intelligent tutoring system for deaf learners of written English. In M. Tremaine, E. Cole, & E.D. Mynatt (Eds.), Proceedings of the ACM Conference on assistive technologies (pp. 92–100). Arlington, Virginia, USA: ACM.

Michaud, L.N., McCoy, K.F., & Stark, L.A. (2001). Modeling the acquisition of English: An intelligent CALL approach. In M. Bauer, P.J. Gmytrasiewicz, & J. Vassileva (Eds.), User Modeling 2001: Eighth international conference (pp. 14–23). Berlin, Heidelberg: Springer Verlag.

Mislevy, R. (2006). Cognitive psychology and educational assessment. In R. Brennan (Ed.), *Educational measurement* (pp. 257–306). Phoenix, AZ: Greenwood.

Murphy, M., & McTear, M. (1997). Learner modelling for intelligent CALL. In A. Jameson, C. Paris, & C. Tasso (Eds.), User modeling: Proceedings of the sixth international conference (pp. 301–312). Wien, New York: Springer Verlag. Retrieved September 1, 2008, from http://um.org/um97/abstracts/MurphyM.htm

Nagata, N. (2002). BANZAI: An application of natural language processing to web based language learning. *CALICO Journal*, *19*(3), 583–599. Retrieved September 1, 2008, from http://www.usfca.edu/japanese/CALICO02.pdf

Nunan, D. (1989). *Designing tasks for the communicative classroom*. Cambridge, UK: Cambridge University Press.

Odlin, T. (2003). Cross-linguistic influence. In C. Doughty, & M. Long (Eds.), *Handbook on second language acquisition* (pp. 436–486). Oxford: Blackwell.

O'Malley, M., & Chamot, A. (1990). *Learning strategies in second language acquisition*. New York: Cambridge University Press.

Richards, J., Plat, J., & Weber, H. (1985). *Longman dictionary of applied linguistics*. London: Longman.

Schwartz, B., & Sprouse, R. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, *12*(1), 40–72.

Skehan, P. (1996). Second language acquisition research and foreign language instruction. In J. Willis, & D. Willis (Eds.), *Challenge and change in language teaching*. Oxford: Heinmann.