Computer Assisted Language Learning

Contents and author index, Volume 21, 2008

EDITOR
Jozef Colpaert
Institute for Education and Information Sciences, University of Antwerp, Vezuwsstraat 35, 2000, Antwerp, Belgium.

HONORARY ASSOCIATE EDITOR
Keith Cameron
Villa Effra, Grasse, France.

ASSOCIATE EDITORS
Wilfried Decoo, Brigham Young University, Provo, UT, USA.
Philip Hulhoud, Language Center and Department of Linguistics, Stanford University, Stanford, CA, USA.
Michael Levy, School of Languages and Linguistics, Griffith University, Brisbane, Queensland, Australia.

EDITORIAL BOARD
Lyle Bachman, Department of Applied Linguistics & TESL, University of California, USA; Emerita Bañados, Universidad de Concepción, Chile; Jean-Claude Bertin, Faculté des Affaires Internationales, l'Université du Havre, France; Françoise Bin, School of Applied Language and Intercultural Studies, Dublin City University, Ireland; Chris Bowser, School of Computing & Technology, David Goldman Informatics Centre, UK; Claire Bradin Siskin, Robert Henderson Language Media Center, University of Pittsburgh, USA; Carol Chapelle, TESL/Applied Linguistics, Iowa State University, USA; Howard Hao-Jan Chen, Department of English, National Taiwan Normal University, Taiwan; Deodry Chen, Department of Germanic, Slavic and Semitic Studies, University of California at Santa Barbara, USA; Uchi Felix, Research Centre for New Media in Language Learning, Monash University, Australia; Nina Garrett, Center for Language Study, Yale University, USA; Ana Gimeno-Sanz, Departamento de Lingüística Aplicada, Universidad Politécnica de Valencia, Spain; Dominique Himard, Humanities, Arts and Languages Department, London Metropolitan University, UK; Delph Hoven, Centre for Distance Education, Athabasca University, Alberta, Canada; Jan Houtgraaf, Faculty of Humanities, University of Amsterdam, The Netherlands; Joa Jamieson, Department of English, Northern Arizona University, USA; Derek Lewis, School of Modern Languages, University of Exeter, UK; Mathias Schulze, Department of Germanic and Slavic Studies, University of Waterloo, Canada; Lesley Shield, Department of Languages, Open University, UK; Vance Stevens, Petroleum Institute, United Arab Emirates; Glenn Stockwell, School of Law, Waseda University, Japan; Naoyuki Tokuda, SunFlare Research & Development Center, Tokyo, Japan; Monica Ward, School of Computing, Dublin City University, Ireland.

© Copyright 2008 by Taylor & Francis.
From recording linguistic competence to supporting inferences about language acquisition in context

Luiz Amaral* and Detmar Meurersb

*Department of Languages Literatures and Cultures, University of Massachusetts, Amherst, USA; bDepartment 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 TÁGARELA 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

ISBN 0958-4221 print ISSN 1744-3210 online
© 2008 Taylor & Francis
DOI: 10.1080/0958822002343454
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., Heif, 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., Heif, 2003, 2004; and the articles in this issue). The student model of the E-Tutor keeps track of the 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 for each grammar skill. To classify a student’s knowledge, the numeric performance scores for each grammar skill are grouped into three levels: beginner, intermediate, and advanced. The performance scores are again grouped into three levels: beginner, intermediate, and advanced. 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 a grammatical basis. 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 and ‘unacquired’. The knowledge units of the SLALOM architecture are grammatical 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.,

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 for ‘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’s knowledge 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 tense/aspect 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, Patl 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 commitment, 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. 59) 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, filling 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 Module 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 citecitation, contractions and abbreviations.

The full-form lookup then returns multiple analyses based on the CURUPIRA lexicon (Martins, Hasegawa, & Gracas 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.](image-url)
the multiple lexical analyses based on the local context. Complementing these local disambiguation rules, a simple bottom-up chart parser using handwritten 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 sentence answer and some 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 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 basis of a feedback mechanism) how to generate the appropriate 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 activity based on its course number and the module/lesson number. We take the progression of the course material to be an indication of the complexity of the linguistic forms necessary to perform specific tasks. 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 productions 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 submodules use 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.

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 discuss the benefits of the extended student model in the next section.

- **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, number, case, 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.

- **LI Transfer**: The last repository of the student model is dedicated to LI transfer errors. The LI 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., Olin, 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 used most frequently is to provide feedback 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 two next 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 tasks 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 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
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 LI transfer. The extended learner model reflects current theories of second language acquisition and allows the system to learn from the learner input as part of meaningful language use. We have thus 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 case advantages of a system with access to such extended student models. Once the extended learner model is fully realized, we intend to focus 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 (2007a), 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 Heft, Donna Long, Kathy McCoy, Bob Mislevy, Scott Payne, Mathias Schulze and the reviewers of User Modeling 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/wert). To foster interdisciplinary collaboration as a crucial component of sustainable ICALL research, he coordinates several projects on the interfaces of ICALL (http://purl.org/net/icall) and the ‘Automatic Analysis of Learner Language’ (http://purl.org/net/calico-workshop).

References


The Sentence Fairy: a natural-language generation system to support children’s essay writing

Karin Harbusch*, Gergana Itsova, Ulrich Koch and Christine Kühner

Computer Science Department, University of Koblenz-Landau, Koblenz, Germany

(Received 1 February 2008; final version received 8 June 2008)

We built an NLP system implementing a ‘virtual writing conference’ for elementary-school children, with German as the target language. Currently, state-of-the-art computer support for writing tasks is restricted to multiple-choice questions or quizzes because automatic parsing of the often ambiguous and fragmentary texts produced by pupils presents unsurmountable problems. Here, we follow a different course by deploying natural language generation technology to evaluate and improve the grammatical quality of student output. Based on an abstract representation of the story under construction, all paraphrases of simple and combined clauses are generated fully automatically. From this source, the system produces exercises enabling the pupils to improve their sentences. We apply parsing technology only in the teacher mode, where new stories are entered into the system in a simple manner. The system we describe here is a prototype including a full exercise generation mode and a rudimentary teacher mode.

Keywords: Intelligent CALL (ICALL); natural language generation (NLG); German; essay writing; elementary school

1. Motivation

German elementary school children learn essay and story writing in a holistic manner. Typically, they are presented with a series of about three to five pictures and/or words and then have to produce an interesting short story of about 10 sentences. In a writing conference (Graves, 1983), the whole class evaluates one such essay and discusses techniques to improve the text in a hands-on manner. The Sentence Fairy (Satzfee) is a ‘virtual writing conference’ that aims to improve essay-writing skills of German elementary school pupils through short exercises, with well targeted syntactic feedback that is produced automatically. The system does not aim at stimulating the creativity needed to write interesting stories. This aspect is left to the teacher, who can enter new stories into the system via the teacher mode. The software uses existing natural-language processing components, namely a syntactic/semantic parser, and a paraphrase generator based on the Performance Grammar (PG) formalism (Kempen & Harbusch, 2002). Syntactic structures are enriched with semantic features in the spirit of Minimal Recursion Semantics (MRS) (Copestake, Flickinger, Sag, & Pollard, 2005).

*Corresponding author. Email: harbusch@uni-koblenz.de

ISSN 0958-8221 print/ISSN 1744-3210 online © 2008 Taylor & Francis DOI: 10.1080/09588220802054496
http://www.informaworld.com