ICT for Language Learning
International Conference
"ICT for Language Learning" 6th edition

# Recycling Learner Data to Construct Data-driven Learning Tools: Practical Applications for the Learner, the Instructor, and the Researcher

[1]Trude Heift, [2]Catherine Caws
[1]Simon Fraser University, [2]University of Victoria (Canada)
heift@sfu.ca, ccaws@uvic.ca

## 1. Introduction

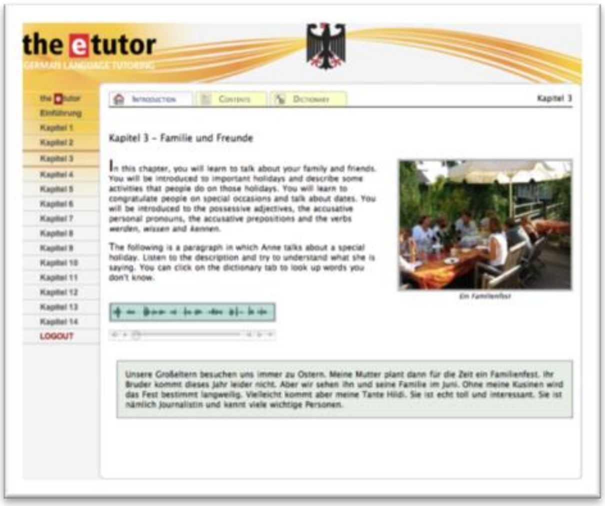
Data collections from learner-computer interactions are commonly used to determine learning outcomes and/or improve software features that are designed to enhance the learning experience for specific learning activities. For this reason, a holistic and cyclical approach to software engineering is generally preferred (see [3], [9]) because each and every stage during the lifecycle delivers output that serves as input for the subsequent stage. Accordingly, these data collections and analyses of learner data have provided feedback and recommendations on the efficacy of certain learning tools as well as their technical usability. However, a much less explored application of learner data collections and analyses has emphasized the recycling of learner data to expand an existing learning environment (e.g., Tutorial CALL) to include data-driven learning (DDL), especially whereby the corpus used for DDL has been constructed from authentic learner submissions of the very same learning environment.

DDL has been explored under various aspects by generally emphasizing the ways in which it can facilitate the implementation of a methodology for language learning that focuses on authenticity in contents, context, and task [10], commonly achieved with a corpus of L1 or L2 data. This article focuses on the recycled data of *E-Tutor*, a comprehensive Intelligent Computer-assisted Language Learning (ICALL) environment for L2 German, and discusses the multiple DDL applications of its collected user data with respect to the learner, the instructor, and the researcher. *E-Tutor* contains a learner corpus that was constructed from user data collected from approximately 5000 students over five years. Accordingly, *E-Tutor* follows a cyclical process of development, implementation and evaluation (see [4]) to inform language teaching pedagogy and provide system enhancements generated by the outcomes of data collections, in particular, with regard to interface design, error analysis, help options, system feedback, and DDL.

In the following, we first introduce *E-Tutor* by describing its system functionality as well as the learner corpus that we constructed from recycled learner data. Next, we discuss the learner corpus with respect to its applications and uses for the learner, instructor, and researcher. The article concludes with a discussion of opportunities for further system and extensions and research.

## 2. *E-Tutor* content and system functionality

*E-Tutor* is an ICALL system for beginner and intermediate learners of German that covers learning content distributed over a total of fifteen chapters. Each chapter begins with an introductory text (e.g., story, dialogue) that highlights the focus of the chapter. For instance, Figure 1 shows the introductory page of Chapter 3 of *E-Tutor*, which centers around the topic of family and friends. From here, learners then have access to the chapter contents (Contents tab) as well as a bilingual dictionary that contains approximately 20,000 entries (Dictionary tab).



Kapitel 3 – Familie und Freunde

In this chapter, you will learn to talk about your family and friends. You will be introduced to important holidays and describe some activities that people do on those holidays. You will learn to congratulate people on special occasions and talk about dates. You will be introduced to the possessive adjectives, the accusative personal pronouns, the accusative prepositions and the verbs *werden, wissen* and *kennen*.

The following is a paragraph in which Anne talks about a special holiday. Listen to the description and try to understand what she is saying. You can click on the dictionary tab to look up words you don't know.

Unsere Großeltern besuchen uns immer zu Ostern. Meine Mutter plant dann für die Zeit ein Familienfest. Ihr Bruder kommt dieses Jahr leider nicht. Aber wir sehen ihn und seine Familie im Juni. Ohne meine Kusinen wird das Fest bestimmt langweilig. Vielleicht kommt aber meine Tante Hildi. Sie ist echt toll und interessant. Sie ist nämlich Journalistin und kennt viele wichtige Personen.

Figure 1: Introductory screen of a chapter

Each chapter offers different learning activities that allow students to practice chapter-related vocabulary and grammar. In addition, there are learning activities for pronunciation, listening and reading comprehension, culture and writing. There are currently ten activity types implemented in the system (e.g., sentence building, reading comprehension, essay) in addition to an introductory unit on pronunciation. For example, in the sentence-building activity, students are asked to construct a sentence from grammatical cues and words that are provided in their base forms. For the reading comprehension activity, students study a chapter-related text and answer comprehension questions. For the essay, students write a minimum of fifty words at the introductory level. As with the other activity types, the essay topics vary by chapter and are always closely related to the content, vocabulary and grammar taught in each chapter. Students have also access to web links that relate to the current chapter content, authentic pictures, grammar notes, statistics on system use and user performance, as well as learner progress reports.

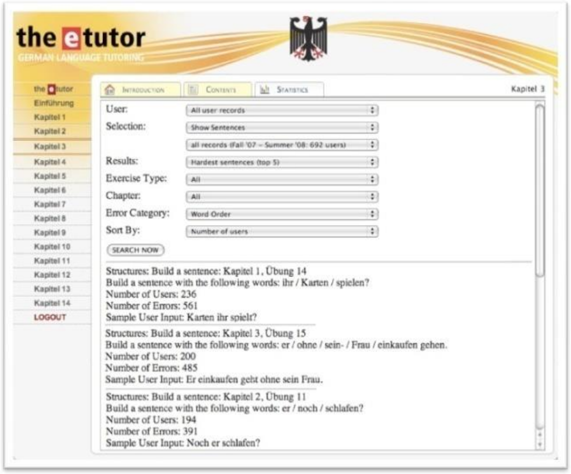
From a computational point of view, *E-Tutor* is an ICALL system for L2 German that integrates Natural Language Processing (NLP) and Artificial Intelligence (AI) modeling into CALL. NLP techniques model 'understanding' of human language by computer, while AI techniques can be used to model the individualized learning experience; the process aims at learning programs that come closer to natural language interaction between humans than has been the case in traditional CALL. The NLP component performs a linguistic analysis of learner input by checking for correct syntax, morphology, and to a lesser extent, semantics (see [7]). Once it has identified the correct and incorrect structures in the learner's input, the system obtains learner and task-specific information from the learner model for feedback generation. The learner model provides a dynamic assessment of each learner by considering past and current learner performance and behavior in relationship to a particular learning activity (see also [1], [8]). Thus the system's interaction with each student is individualized as to the kinds of errors in the student input as well as the ways they are communicated to the learner.

*E-Tutor* has been used for over a decade at several North American universities and learner data gathered by *E-Tutor* have been analyzed to answer specific research questions that assess learning outcomes as well as system features and functionality. But in addition to these smaller-scale data collections and analyses, millions of user submissions from a variety of learning tasks in *E-Tutor* were collected over a time period of five years and from those a common learner corpus was constructed. While this tool is fairly easily constructed due to the underlying NLP analysis of *E-Tutor* and the user log that the system constructs from all user inputs, it has multiple applications as they relate to the learner, the teacher, and the researcher.

## 3. Applications for the learner, instructor, and researcher

The learner corpus that we constructed from approximately 5000 previous system users allows learners to examine interlanguage or task-specific phenomena and the benefits in this respect have been well documented (see e.g., [2], [5], [6]). In addition and given that the system also collects system submissions for each student while using the system concurrently, learners can compare their own data with that of all previous users and determine how and in what ways their interlanguage differs from that of other users. Accordingly, the data that the system collects during system use is continuously being recycled to provide new learning tools for the learner. Figure 2 displays the interface of the searchable learner corpus.

As indicated in Figure 2, learners can choose between the common learner corpus that combines all user submissions or they can limit the search to only their own corpus. If they select to search all user records, they can also specify the time period for which they want to execute their search (e.g., all five years or one particular year). In addition, users can filter their search by exercise type, chapter and error category and thus inspect what kind of errors commonly occurred in which chapters and activity types. Finally, learners can also choose between the hardest and easiest sentences. The hardest sentences are those where either most learners committed an error (sort by "number of users") or the sentences that contained most errors (sort by "number of errors"). Conversely, the easiest sentences are those that were submitted correctly by most learners (sort by "number of users") or the sentences that contained the fewest errors in total (sort by "number of errors"). The lower half of Figure 10 displays the search results, that is, in this instance, the exercise type and instruction, the number of users and errors, and the sentence that, according to the parameters selected was answered incorrectly by the students.



User: All user records
Selection: Show Sentences
all records (Fall '07 – Summer '08: 692 users)
Results: Hardest sentences (top 5)
Exercise Type: All
Chapter: All
Error Category: Word Order
Sort By: Number of users

Structures: Build a sentence: Kapitel 1, Übung 14
Build a sentence with the following words: ihr / Karten / spielen?
Number of Users: 236
Number of Errors: 561
Sample User Input: Karten ihr spielt?

Structures: Build a sentence: Kapitel 3, Übung 15
Build a sentence with the following words: er / ohne / sein- / Frau / einkaufen gehen.
Number of Users: 200
Number of Errors: 485
Sample User Input: Er einkaufen geht ohne sein Frau.

Structures: Build a sentence: Kapitel 2, Übung 11
Build a sentence with the following words: er / noch / schlafen?
Number of Users: 194
Number of Errors: 391
Sample User Input: Noch er schlafen?

Figure 2: Learner language

Accordingly, the DDL tool of *E-Tutor* provides multiple learning opportunities for users. Most importantly, learners can examine interlanguage or task-specific phenomena by contrasting their own data against that of all users, and they can determine and analyse the variations between their interlanguage and that of other users.

In addition to these learner applications of the corpus, the learner corpus also provides opportunities for the language instructor. For instance, the learner corpus allows language instructors to examine the design of language learning material in that the data as well as their statistics reveal much about the degree of difficulty of each exercise, in particular, if they have been tested on approximately 5,000 learners. Accordingly, instructors can readily search which exercise type caused the least or the most difficulties and even more specifically, which target words as well as grammatical constructions triggered the most errors. As a follow-up to the regular classroom activities, the instructor can then focus on the problematic areas as revealed by the learner corpus. Accordingly, the corpus serves for material selection with regards to language exploration and remediation which also helps students with the meta-cognitive analysis of the L2.

Finally, these large data collections also allow for investigations of a wide range of additional research topics to be explored by the researcher. For instance, the learner corpus provides rich data for interlanguage studies given that the data can also be correlated to leaner variables such as L1 background and L2 exposure, among others. Moreover, given that the system also logs the learner interaction with the corpus, researchers can investigate in what ways learners make use of their own recycled data.

## 4. Conclusion

In conclusion, DDL can facilitate the implementation of a methodology for language learning that focuses on authenticity in contents, context, and task. By employing a holistic and cyclical approach to software development, a learner corpus can be constructed by making use of the learners' own data to provide an authentic learning environment that can be explored by learners, instructors, and researchers alike. By collecting new data on learner use of the corpus, the learner-computer interaction can be enhanced by new insights gained from new data collections thus highlighting the benefits of a cyclical approach to software engineering.

## References

[1] Amaral, L., & Meurers, D. (2007). Conceptualizing Student Models for ICALL. Paper presented at: *User Modeling 2007: 11th International Conference*. June 25-29, 2007, Corfu, Greece.

[2] Braun, S. (2005). From Pedagogically Relevant Corpora to Authentic Language Learning Contents. *ReCALL, 17*(1), 47-64.

[3] Caws, C. (2013). Evaluating a web-based video corpus through an analysis of user interactions. *ReCALL, 25*(1), 85-104.

[4] Colpaert, J. (2006). Pedagogy-driven design for online language teaching and learning. *CALICO Journal, 23*(3), 477.

[5] Gaskell, D., & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System, 32*, 301-319.

[6] Granger, S. (2003). Error-tagged Learner Corpora and CALL: A Promising Synergy. *CALICO Journal, 20*(3), 465-480.

[7] Heift, T. (2010). Developing an Intelligent Language Tutor. *CALICO*, 27(3), pp. 443-459.

[8] Heift, T., & Schulze, M. (2007). *Errors and Intelligence in CALL: Parsers and Pedagogues.* New York: Routledge.

[9] Hubbard, P. 2011. Evaluation of Courseware and Websites. In L. Ducate and N. Arnold (Eds.) *Present and Future Perspectives of CALL: From Theory and Research to New Directions in Foreign Language Teaching*, Second Edition. San Marcos, TX: CALICO.

[10] Van Lier, L. (1996). *Interaction in the Language Curriculum: Awareness, Autonomy, and Authenticity.* London: Longman.