Learning the Learner

User Modeling in Intelligent Computer Assisted Language Learning Systems

David Alfter
Språkbanken
University of Gothenburg, Department of Swedish, Box 200
405 30 Gothenburg
david.alfter@svenska.gu.se

ABSTRACT

The concept of user modeling, and more specifically learner modeling, has been known for quite a long time, but its implementation into language learning systems is still rather uncommon. The current work proposes to extend an existing experimental language learning platform with user modeling capabilities in order to offer a more personalized experience to the different users.

Keywords

user modeling; language learning; ICALL

1. INTRODUCTION

In order to overcome the limitations of traditional computer assisted language learning (CALL) systems, research has investigated the use of natural language processing (NLP) tools in language learning, creating the discipline of Intelligent CALL (ICALL) [15].

The problem is that language learning platforms mostly present their knowledge in the same form for every learner. Every learner, however, has different needs [6]. One learner might prefer written texts, another one might prefer videos, yet another one might prefer songs. The most learner-friendly solution would be to implement a dedicated learner module for every learner. However, this is not feasible [6].

If we can offer a unified learning and testing environment that adapts to different learners, we can narrow the gap between the expected 'one learner - one platform' thinking and the actual realization. Thus, learner modeling is important if we want to individualize language learning for a given learner [14]. This way, we can select the most appropriate pedagogical solutions for a learner given the information that the learner model gives us [14]. This is also very important if we want to give the learner personalized feedback [14].

2. MOTIVATION

While there is ongoing research at the international level, Swedish ICALL systems are rare, despite the availability of the necessary resources [15]. The present work aims at improving an existing language learning platform for learners of Swedish by adding user modeling capabilities to the system.

3. RELATED WORK

3.1 Accelerated Learning

Many different factors, such as social and cognitive factors, affect language learning [11]. Among those factors, the native language of the language learner is a frequent cause of L2 errors [11, 17, 5]. This phenomenon is known as language transfer [11].

This also means that it is possible to reason about the L1 from an L2 production. [17] have shown that syntax errors in L2 can be used to infer the native language of the learner with an accuracy of 71.71%; [8] even reach 84.3% accuracy. However, generalizing about speakers of the same L1 yields suboptimal results [11]. Indeed, not every speaker of a given L1 is influenced in the same way by the L1 and speakers of different L1s may commit the same error for different reasons [11].

As [5] points out, the mother tongue influences second language acquisition, but not necessarily in a bad way. There are inhibitory, but also facilitative interferences [5].

Depending on the language background of a learner, different concepts in a target language appear more or less difficult. If the learner already knows a language that exhibits 'gender', for example German or French, 'gender' in Swedish will be less difficult a concept to grasp. If, however, the learner does not know a 'gender' language, the concept will be much harder to understand. Similarly, if the learner knows only non-inflecting languages, inflection will prove more difficult to learn.

If we can model a learner using this dimension, we might be able to accelerate language learning for learners depending on their language background by introducing *learner-easy* topics first and putting more emphasis on *learner-hard* topics, where *learner-easy* topics are topics that the learner already knows from previous languages and *learner-hard* topics are topics that the learner has not encountered in any of the previously learned languages.

3.2 Individuals and Groups

In order to model a learner, it is necessary to collect data from the learner [6]. This can be personal data and/or data about the user behavior on the language learning application [6].

As it is impractical to model and track each learner separately, [6] introduced the term 'persona'. A persona captures and clusters similarities and differences among learners [6]. It is probable that similar users have similar needs, so instead of addressing each user individually, personas allow an

ICALL platform to remain flexible while being individualized for each learner.

A similar concept, named 'performance profiles' has been proposed by [10]. In a previous work, [9] found that learners within the same proficiency group tend to commit the same errors. The introduced performance profiles are used to capture statistically significant differences in different learners' grammar [10].

Our learner grouping builds on these ideas. The intuition is that there are sub-groups of learners that commit the same errors and have similar needs concerning feedback. These group patterns should be recognized or learned using statistical models, rather than be based on static demographic data of the learner. Indeed, it has been shown that the influence of culture and mother tongue (L1) on second language (L2) learning is not the same for all speakers of the same L1 and that predictive error models based on generalized stereotypes have low accuracy [11]. By calculating the similarity of learners on the basis of the committed errors, the approach is thought to be more unbiased and robust than stereotyping or traditional personas.

On the other hand, by logging the progress of an individual learner, it is possible to evaluate and re-evaluate the language proficiency over time and adapt the exercises accordingly. Another advantage is that individual learner variables can be used to discover correlations between different variables and language learning. However, it is not yet clear whether there are significant correlations. For example, [6], among others, has found that gender does not have an influence on language learning. Thus, we must first answer the question of which user variables to collect, as the choice of variables influences the way in which we can model learners.

3.3 Error Classification

Like learners, errors can be classified along different dimensions. [5] broadly distinguishes between grammatical, phonologically induced, lexical and discourse errors. [16] propose a different system with fine-grained and domain-adapted distinctions in order to provide meaningful feedback. [13] distinguish between lexical errors and errors in tense, mood, agreement and conjugation among others. It should be noted that, to a certain degree, the chosen classification always depends on the task and on the language at hand.

Another important distinction is often made between 'mistakes' and 'errors' [14, 16]. In this parlance and in Chomskyan terms, errors are competence-based and mistakes are performance-based [14]. This means that errors are committed because the learner simply does not know the correct answer whereas mistakes are made despite the learner knowing the correct form [14].

By classifying errors into different categories, we can also cluster learners by error category. The intuition behind this approach is that users who commit similar errors (i.e. they have a similar error profile) will benefit from similar remedial actions.

3.4 Feedback

Feedback is an important part of learning [11]. The effectiveness of feedback depends on many variables, e.g. the clarity of the feedback, the way it is given, or the student with all his mental and physical states [11].

Feedback can be positive or negative; negative feedback is used to correct errors and to prevent the learner from repeating that error [4]. Positive feedback is used to encourage the learner, to reinforce correct knowledge and to integrate new knowledge that might have resulted from tentative or random answering [4]. Most learning systems use negative feedback, as it is easier to implement [4].

However, feedback also has an influence on the attitude and motivation of the learner [11]. Feedback, especially negative feedback, can be perceived as threatening to the self-esteem and confidence [2], and might lead to a decrease in motivation. Positive feedback, on the other hand, is seen as a motivational technique [4] and students given positive feedback perform better than students given negative feedback [1].

Also, in order to be effective, feedback must not be too long or too short [11]. [13] have found a correlation between a learner's language proficiency and the optimal amount of feedback, i.e. the more proficient a learner is, the less feedback is needed.

3.5 Cold Start Problem

One problem in user modeling is encountered at the very beginning when the system does not have any knowledge about the user. This is called the 'cold start problem' [12]. As we cannot wait until we have full knowledge about the users before we start reasoning about them, we have to find some sensible way of dealing with this situation [3]. One possible solution is to use 'stereotypes' [7, 3, 9]. Stereotypes group together users with similar characteristics [7, 3]. After the initial 'cold start boundary' has been overcome, regular user models have been shown to perform better and should replace or augment the initial models [12]. Other solutions include setting default rules that are applied unless we have information about the user that invalidates the default rules and negation as failure [3].

4. PROGRESS MADE TO DATE

As the PhD project is in its incipient stages, most progress that has been made has been of theoretical nature.

5. PROPOSED APPROACH

First of all, we have to decide which user variables and which user data to collect. Depending on the chosen variables and data, we then define different tasks for the language learners. These tasks will be implemented into an existing experimental platform that will automatically collect all the required data. We then ask students to work on the tasks. The collected data will be evaluated and used to improve the user modeling capabilities of the system. Finally, we ask students and teachers to use the platform again so as to gain data which will be evaluated in order to improve the platform further.

5.1 Data

Learner data can be divided into several groups. One group concerns personal information such as sex, highest educational level, native language(s). Another group concerns implicitly gained knowledge as the learner uses the learning platform, such as time spent on exercises, number of tries, types of errors. We plan on collecting both types of information.

User data will be handled anonymously and personal data will not be used for identification purposes.

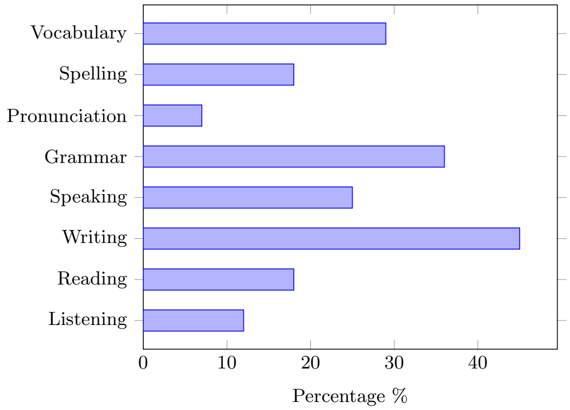

Figure 1: Distribution of practiced skills

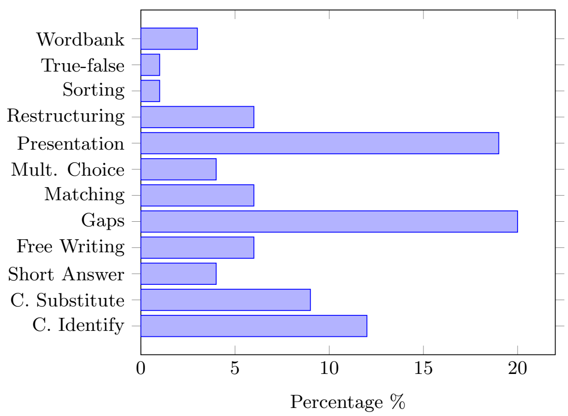

Figure 2: Distribution of exercise types

5.2 Experimental Design

When all the technical details have been cleared and all exercises and data collection services have been set up, there will be a first case study with learners of Swedish as a second language. The goal of this case study is to collect data which will be evaluated and the evaluation of which will form the basis for the subsequent user models.

In a first step, we will target intermediary language learners, i.e. learners who have already acquired the basics of the language, but are not yet independent language users. Preliminary analyses using course book corpora have shown that at this level, most course books concentrate on writing (see figure 1), and the most common exercise type consists of gapped exercises (see figure 2).

Given these findings and the fact that we already have a working vocabulary exercise generator, we will first concentrate on vocabulary exercises.

Vocabulary exercises can be used to broadly assess the language knowledge of a learner by comparing the target vocabulary and a frequency word list; more frequent words would be learned first and more rare words at a later stage. Knowing a word at a certain position in the list would presume knowledge most words more frequent than that word.

As we want to follow the course book pedagogical path, we choose gap tests (cloze tasks). The user is presented with a sentence or a paragraph of coherent text, and some of the words are taken out and have to be filled in by the learner. The target vocabulary will be based on the learner's knowledge.

Cloze tasks can be used not only to assess vocabulary knowledge, but also for checking agreement or inflectional knowledge. Cloze tasks can also be used to determine collocational knowledge by selecting sentences containing strong collocations and specifically gapping one of the involved words.

We also plan on using cloze tasks, among other tasks, for a language diagnostic test. The test will be made up of sets of four to five unconnected sentences that have been chosen based on the learner's language level. All of the sentences in a set will contain the same target word, but not necessarily in the same surface form; the target word might be in the singular in sentence one and in the plural in sentence two, or it might be a verb used in different tenses. At least one of the sentences will contain a strong collocation containing the target word. The learner will be told that all of the gaps contain the same word, but not necessarily in the same inflection; they should write the word that they think fits best. The learner will then be presented one sentence after another, with the next sentence only appearing after the current gap has been filled. It will not be possible to go back to previous gaps and change the input. This approach tests several skills simultaneously, and at the same time doesn't take too long to complete. We hope to be able to classify learners and assess their lexical language knowledge at least broadly using this approach.

At a later stage, grammar exercises will be introduced. These exercises allow for a more fine-grained error classification, but they also require more extensive evaluation and more sophisticated analysis components.

The next step concerns user models properly speaking. First, a theoretical user model will have to be created. This model should indicate how different user variables should be linked and evaluated to arrive at either a quantitative or a qualitative representation of the learner and the learner's progress. This model will then also be implemented in the experimental platform.

At this stage, a second case study will be organized. This study, in contrast to the first study, also takes teachers of Swedish as a second language into account. The aim of the second study is to confirm the findings of the first study, but also to take teacher feedback into account.

We hope to arrive at a mature system that will be able to track user's progress, adapt to the user, recommend learning paths based on data analysis and user models and give feedback in a personalized manner.

5.3 Evaluation Criteria

After data collection, the collected data will have to be evaluated. Data from the first case study will serve as a basis for preliminary user models. The result of this evaluation should give insights into which user variables are most important and how the user variables can be used to model learners with regard especially to the errors and mistakes they made.

The second case study evaluation will concern the "effectiveness" of the preliminary user models. Furthermore, as the second case study also takes teacher feedback into account, its evaluation could serve as a basis for automated feedback generation.

6. FURTHER RESEARCH

Further research should cover automatic individualized feedback generation in more depth. Feedback generation is another very big field which can be regarded both as distinct from user modeling, as well as being a part of user modeling. If we adopt the latter view, feedback can be individualized as well.

The final aim of the experimental language learning and studying platform is to cater to different groups of people, offering a comprehensive online resource that connects researchers, linguists, learners of Swedish as a second language and teachers of Swedish as a second language.

Acknowledgment

I would like to thank my supervisors Lars Borin and Elena Volodina for their helpful insights and guidance.

7. REFERENCES

[1] Corrigan-Halpern, A. *Feedback in complex learning: Considering the relationship between utility and processing demands*. PhD thesis, University of Illinois, Chicago, 2006.

[2] Fatehi, M., and Akbari, O. An Experimental Analysis of Errors in Light of Language Learning and Language Use and the Role of Executing Involvement to Increase Motivation in the English Language Classroom. *International Letters of Social and Humanistic Sciences 51* (2015), 82–88.

[3] Finin, T., and Drager, D. GUMS: A General User Modeling System. In *Proceedings of the workshop on Strategic computing natural language* (1986), Association for Computational Linguistics, pp. 224–230.

[4] Fossati, D. The role of positive feedback in intelligent tutoring systems. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Student Research Workshop* (2008), Association for Computational Linguistics, pp. 31–36.

[5] Hanafi, A. The second language influence on foreign language learners'errors: The case of the french language for algerian students learning english as a foreign language. *European Scientific Journal 10*, 10 (2014).

[6] Heift, T. Learner Personas in CALL. *CALICO Journal 25*, 1 (2007), 1–10.

[7] Kass, R., and Finin, T. Modeling the User in Natural Language Systems. *Computational Linguistics 14*, 3 (1988), 5–22.

[8] Malmasi, S., Tetreault, J., and Dras, M. Oracle and Human Baselines for Native Language Identification. In *Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications* (2015), pp. 172–178.

[9] Michaud, L. N., and McCoy, K. F. Modeling User Language Proficiency in a Writing Tutor for Deaf Learners of English. In *Proceedings of a Symposium on Computer Mediated Language Assessment and Evaluation in Natural Language Processing* (1999), Association for Computational Linguistics, pp. 47–54.

[10] Michaud, L. N., and McCoy, K. F. Error Profiling: Toward a Model of English Acquisition for Deaf Learners. In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics* (2001), Association for Computational Linguistics, pp. 394–401.

[11] Myles, J. Second language writing and research: The writing process and error analysis in student texts. *The Electronic Journal for English as a Second Language 6*, 2 (2002), 1–20.

[12] Sahebi, S., and Brusilovsky, P. Cross-domain collaborative recommendation in a cold-start context: The impact of user profile size on the quality of recommendation. In *User Modeling, Adaptation, and Personalization*. Springer, 2013, pp. 289–295.

[13] Shaalan, K. F., and Magdy, M. Adaptive Feedback Message Generation for Second Language Learners of Arabic. In *Proceedings of the 8th International Conference on Recent Advances in Natural Language Processing* (2011), pp. 752–757.

[14] Thouësny, S., and Blin, F. Modeling language learners' knowledge: What information can be inferred from learners' free written texts? *WorldCALL: International Perspectives on Computer-Assisted Language Learning. Routledge, New York* (2010).

[15] Volodina, E., and Borin, L. Developing an open-source web-based exercise generator for Swedish. In *EUROCALL Conference* (2012), pp. 307–313.

[16] Volodina, E., and Pijetlovic, D. Lark Trills for Language Drills: Text-to-speech technology for language learners. In *Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications* (2015), pp. 107–117.

[17] Wong, S.-M. J., and Dras, M. Contrastive analysis and native language identification. In *Proceedings of the Australasian Language Technology Association Workshop* (2009), pp. 53–61.