Learners’ situated motivation in oral grammar practice with an ASR-enabled CALL system

Stephen Bodnar, Bart Penning de Vries, Catia Cucchiarini, Helmer Strik, Roeland van Hout

Centre for Language and Speech Technology, Radboud University Nijmegen, the Netherlands
{B.Penningdevries, S.Bodnar, C.Cucchiarini, W.Strik, R.vanHout}@let.ru.nl

Abstract

Advances in speech recognition (ASR) technology have resulted in computer applications that provide compelling forms of speaking practice to learners of a second language (L2). Evaluation of such applications typically does not include an analysis of how learners’ situated motivational states fluctuate during language practice. In connection with recent developments in L2 motivation theory, this paper investigates situated learner motivation in practice with an ASR-enabled system. An experiment was conducted in which our system provided oral grammar practice for Dutch L2 under two learning conditions: with and without corrective feedback. We report on learners’ motivational experiences by triangulating an analysis of 1) motivational trajectories from a periodic motivation questionnaire, 2) post-practice reflective questionnaires, and 3) behavioural log data recorded by the system during practice. Our analysis shows that learners maintained positive attitudes towards the system throughout practice and became increasingly confident over time.

Index Terms: CALL, motivation, ASR, corrective feedback, grammar

1. Introduction

Advances in automatic speech recognition (ASR) technology have resulted in computer applications that provide interactive speaking practice to learners of a second language (L2). Different types of interaction are possible, including simulated conversational exchanges and pronunciation training. Typically, evaluation of such applications has concentrated on the performance of the technology, or the impact on L2 proficiency, rather than on learner motivation. When motivation has been included, it has typically been addressed through pre- and post-training evaluation questionnaires. Thus, little is known about how learners’ situated motivational state fluctuates during practice with ASR-enabled systems.

Connected to this gap in the research is work by [1] which advocates viewing motivation as a learner state that emerges from a series of interactions between learner and context. Furthermore, it has recently been suggested to complement subjective methods with objective records of events that occur during practice [2, p.68].

In this paper we investigate situated learner motivation in practice with an ASR-enabled system that provides oral grammar practice for L2 learners of Dutch. The system can be configured in different ways to allow different forms of learner-system interaction. For instance, the system elicits spoken output from the learners, which later can be evaluated through ASR to provide an overall score. Another configuration option is to employ ASR to provide immediate corrective feedback on each utterance. This makes it possible to investigate situated motivation under different learning conditions. An important issue in second language acquisition (SLA) research is the effect of corrective feedback on L2 development. Corrective feedback can affect learner proficiency, but is also related to learner motivation [3]. Our system employs a logging module that records learner and system behaviour with temporal precision and consistency; in our view, this practice environment presents compelling opportunities to study situated motivation in a way that closely corresponds with current views in the L2 motivation literature. Recently, we conducted an experiment with our system which aimed at investigating situated motivation under different learning conditions, specifically with and without immediate corrective feedback. We report on the motivational experiences of the participants by triangulating an analysis of 1) motivational trajectories from a periodic motivation questionnaire, 2) post-practice reflective questionnaires, and 3) behavioural log data recorded during practice.

2. Situated motivation and Corrective Feedback in L2 learning

In SLA research, motivation is a multi-faceted concept. Early views of motivation took a social-psychological approach where research emphasised the role of learners’ reasons or motivations for learning an L2 [4]. Though considered pioneering for the inclusion of social context in the study of motivation [5, p. 67], later research would criticise this macro perspective, which dominated the study of L2 motivation for many years. A criticism particularly relevant here was that the approach did not
allocate much of a role to motivation in the classroom or other learning contexts [6], in contrast with the views of L2 educators and educational psychology researchers.

More recently, situated perspectives that involve ‘a more fine-tuned and situated analysis of motivation as it operates in actual learning situations’ [5, p.74] have begun to receive attention in the literature. Particularly relevant here are the views of Ushioda, who has described a person-in-context approach to studying situated L2 motivation [2]. Important in this view is the acknowledgement that motivation in a situated context can change over time and that motivation should be studied in relation to interactions or events that take place during language practice. Thus, Ushioda [2] can be seen as calling for fine-grained objective practice data to complement subjective data gathered in situated motivation studies.

In our view, this approach to motivation is well-suited for research into language learning in a computer environment. The reason is that CALL applications can support a variety of learner-system interactions. A second reason is that they have the capability to log in great detail these interactions. Despite these capabilities, few studies have researched situated motivation in a CALL environment. Notable exceptions are work by [7], who studied the effect of a personalisation strategy in an L2 vocabulary tutor, [8], who used computer logs to study motivation in an online language course, and [9], who investigated the automatic detection of learners’ motivational state from computer log files.

As a step towards employing computer capabilities to study situated motivation in a CALL environment, we built and tested a new motivation component in our CALL system. In an experiment with Dutch L2 learners, we tested the use of our new component in the context of an issue that has received considerable attention in the recent literature: the role of corrective feedback in L2 learning. A considerable body of literature has indicated that one of the problems in research on corrective feedback is the impossibility of studying this phenomenon under tightly controlled conditions [10]. Studying learner motivation in relation to corrective feedback is particularly relevant as various researchers have pointed to a possible demotivating effect [11, 12, 13, 14]. Our system makes it possible to create conditions in which corrective feedback is provided instantaneously, systematically and intensively, with opportunities for self-repair on the part of the learner. The new motivation component, together with the logging capabilities of our system, provide very large quantities of subjective and objective data that contribute to our understanding of motivation and corrective feedback in L2 learning.

3. Materials and Methods

This section presents materials used in the experiment: an ASR-based CALL system for practicing grammar in Dutch L2, situated mini-motivation questionnaires, computer practice logs, and a post-practice questionnaire gathering participants’ impressions of the training. A description of the experimental procedure is also included.

3.1. An ASR-based CALL system for Dutch L2 Oral Grammar Practice: GREET

We use ASR in a system that helps learners practice aspects of Dutch word order. A common difficulty encountered by Dutch learners is inverting the position of the subject and verb when required. Our system provides training for this aspect of Dutch with a collection of video clips and question and answer exercises. Learners first watch a video clip and are then quizzed on the contents of the clip by the system. In the quiz, they respond to questions by recording themselves speaking their answer aloud. For experimental purposes, we have built two versions of the system. In the corrective feedback system (CF), ASR is used to provide immediate corrective feedback. The feedback was designed to be of a prompt type (for a detailed description see [15]). First, the system notified the learner that it detected an error by displaying a message, formulated as ‘That was incorrect. Please try again’. Second, the system provided the learner with a hint by incrementally revealing the correct word sequence. In the no-feedback system (NOCF), the system displayed a neutral message notifying the learner that their recording has been saved. Prior to beginning practice, the NOCF group received a message stating that their scores would be evaluated at a later time. All other aspects of practice were identical.

3.2. Mini-Motivation Questionnaires

A key component we added for this experiment is a 3-item ‘mini-motivation’ questionnaire we use to track motivational changes. A screenshot of the questionnaire is shown in Figure 1. The questionnaire consists of three semantic differential scales designed to survey attitude, learners’ general attitudes towards practice with the system, motivation, learners’ motivation level as the desire to continue practice with the system or stop practice and do something else, and self-confidence, as the learner-estimated level of difficulty of future practice with GREET. To aid learners’ understanding of the questionnaire items, English translations of the texts are accessible from an on-screen hyperlink.

3.3. Computer practice logs

The GREET system maintains a detailed log of events that occur during practice. The events recorded by the system include page views, number of video clips viewed, number of questions viewed, time on different types of pages, number of recordings, ASR recognition results, type of feedback returned, and others. When a
learner begins an activity, the system creates a practice session object to store the events that occur. In later analyses these sessions serve as records of the interactions that took place during practice. In this experiment, we focus on the number of video clips viewed, the number of questions completed, and the number of attempts made at answering a question.

3.4. Post-practice questionnaire

We used a post-practice questionnaire to obtain learners’ subjective evaluations of the system. The questionnaire consisted of seven Likert items (using a 5-point scale):

- One item asked learners to rate the efficacy of training with the system.
- One item asked for learners’ opinions on whether they felt their Dutch had improved as a result of practice.
- Three items asked learners to rate the video clips, question exercises and the system as a whole.
- Two items evaluated learners’ difficulty in understanding the Dutch dialogs in the video clips and answering the practice questions.

3.5. Experimental Procedure

In the fall of 2012, we recruited 31 participants from Radboud In’to Languages, the Nijmegen university language centre, for an experiment with our system (for a description of an earlier experiment, see [16]). We focused on students studying at the A1 or A2 level of the Common European Framework (CEF). Participants were randomly assigned to one of two groups: a group who practiced without CF (NOCF group) or a group which practiced with CF (CF group). All other activities were equivalent. Each participant completed two sessions. In session 1, learners logged into the system and completed a background questionnaire and two proficiency tests before beginning practice. In each session they practiced for a total of 45 minutes. Practice was divided into three 15-minute micro sessions (for a total of 6 micro sessions) in which learners practiced spoken Dutch with our system. Before beginning the micro sessions, participants completed the mini-motivation questionnaire for the first time. Then they practiced with the system for 15 minutes, three times in a row in each session, resulting in three micro sessions. At the end of each micro session, they completed the mini-motivation questionnaire. That means that each session has four data points and that we have eight data points overall (two sessions).

4. Results

In our analysis of the situated motivation data, we refer to the collection of motivation ratings recorded by the participant at the beginning (S1-0 for session 1 and S2-0 for session 2) or after finishing the three micro sessions (S1-1 to S1-3 and S2-1 to S2-3). Excluded from the following analyses are the data of three subjects who did not complete all eight mini-motivation quizzes.

4.1. Situated motivation

Differences in situated motivation between the NOCF and CF group were analyzed using a repeated measures ANOVA design. In our tests each micro session is one level in the variable time in the analysis. We tested for effects of group, time (all eight mini quizzes) and the interaction effect between the two. The analysis showed significant effects for time, but not for group or the group-time interaction. We observed a time effect for all three situated motivation items (Attitude: $F(7,189)=2.21$, $p=0.0349$; Motivation: $F(7,189)=7.09$, $p=0.000$; Self-confidence: $F(7,189)=3.76$, $p=0.001$). As there was no group effect (the CF vs the NOCF group) at all, we proceeded with analyzing the changes over time by including session (the first and second session) and micro session (the four mini quizzes with each session) as independent within-subjects variables. Changes in learners levels of attitude, motivation and self-efficacy can be seen in Figure 2.

Analysis of variance for attitude returned a near-significant effect for session ($F(1,28) = 3.332$, $p = .079$), a significant effect for micro session ($F(3, 84) = 3.465$, $p = .020$), and no effect for the interaction between session and micro session ($F < 1$). Although the factor session was not significant, the mean scores in session 2 were higher (4.22) than in session 1 (4.01), indicating that the
scores were not decreasing in any way by doing similar exercises again in the second session. It seems an important conclusion in favour of the system we used. Looking at the micro sessions (using the Sidak post-hoc procedure) we see that the measurements at the start (S1-0, S2-0) have the lowest mean scores (3.83) which differ significantly from micro sessions S2-1 and S2-2, which have higher scores, though not from micro session S2-3 (mean scores respectively 3.83, 4.22, 4.16). Learners seem to enjoy the training sessions more than they expected at the start, both in sessions 1 and 2. It can be taken as an indication that practice contributed to enhancing attitude.

For motivation to practice, both the factor session (F(1,27) = 7.234, p = .012) and the factor micro sessions (F(3,84) = 9.727, p = .000) turn out to be significant, whereas their interaction is just not (F (3,84) = 2.663, p = .053). The session effect suggests that scores in the second session are higher than in the first, a positive outcome implying that practice with the system does not evoke negative experiences. The micro session effect indicates a falling tendency as practice time increases within the same session. This effect is supported by the outcomes of a post-hoc test (Sidak procedure) which points to micro session S1-3 as the most conspicuous one (significantly different from all other micro sessions in session 1, no other pair-wise differences being significant; also significantly different from micro sessions S2-1 and S2-2 in session 2, again no other pair-wise differences being significant). The nearly significant interaction effect seems to point to a stronger falling trend for micro session 3 in session 1. The motivation returns at a higher level again at the beginning of session 2.

For situated confidence, the statistical analysis shows both a session effect (F(1,28) = 4.222, p = .045) and a micro session effect (F(3,84) = 5.9906, p = .001), with interaction between the two not significant (F(3,84) = 1.307, p = .277). The session effect means that the scores in session 2 are higher than in session 1, a result in favor of our system. The post-hoc analysis (Sidak procedure) of the micro session effect points out that micro session S1-0 stands out as the condition with the lowest score within the two sessions (significantly different from all micro sessions in session 2). Micro session S2-3 is significantly different from micro sessions S2-1 and S2-2, suggesting a rising increasing trend. The differences in session 1 are too small to deliver significant post-hoc results). The higher score in session 2 and the rising trend within the same session seem to evidence that the learner gets more confident during practicing.

4.2. Post-practice questionnaire

The values for the CF group were systematically higher for all seven relevant questions with an exception for an item concerning learners’ evaluations of the practice questions (Q13. The questions were . . . (Boring - Nice)) which had identical means (3.8). The overall mean of all seven questions is just not significant for a group effect between the NOCF (mean = 3.74, SD = .693) and CF group (mean = 4.14, SD = .404). It is important to note that both mean scores are high, indicating a positive post-practice evaluation. We found a significant difference in favor of the CF group in a previous experiment [16]. Removing question 13 makes the group difference significant (F (1,27) = 5.153, p = .031). Given the result of the previous experiment our cautious conclusion is that the CF group tend to be more positive than the NOCF group.

4.3. Practice logs

To complement the subjective analysis above, we checked in the practice logs to look at how practice differed for the two groups by conducting a number of repeated ANOVA tests on three measures: the number of video clips watched, questions completed and attempts made per question. Our analysis of practice behaviours showed that there were significant main effects and interactions. The graphs are displayed in Figure 3.

For number of video clips watched, we did not find an effect for group, indicating that both groups watched an equivalent number of videos in each of the micro sessions. We continued the statistical analysis for the effects of session and micro session. There was a clear effect
for session (F(1,28) = 38.628, p = .000), no effect for micro session (F(2,56) = 1.577, p = .216), and a just not significant interaction between session and micro session (F(2,56) = 2.709, p = .075). So, the learners watched more videos in the second session, with perhaps a slight tendency to raise watching frequency during session 2.

For questions completed, a number of trends are visible. Three effects are significant: session ((1,27) = 14.694, p = .001, the interaction between session and group (F(1,27) = 8.819, p = .006, and micro session (F(2,54) = 13.664, p = .000). The last finding, given the trend visible in figure 3, indicates that the number of questions completed by learners increased across practice.

The interaction effect for group by session shows that the increase in questions per session was different for each group. The trend depicted in figure 3 shows a rise in questions completed for the CF group. The difference between both groups can be made visible when the statistical analysis is done for each of the groups separately. In the NOCF group, a significant effect remains for micro session only (F(2,24) = 3.563, p = .04. A post-hoc analysis (Sidak procedure) reveals a difference between micro session S1-1 (lower scores) and micro sessions S1-2 and S1-3 (higher scores). In the CF group, both session (F(1,15) = 27.604, p = .000) and micro session (F(2,30) = 15.212, p = .000) effects are present. The three micro sessions show a constant rising trend, all differences between the micro sessions being statistically significant (Sidak procedure) for the CF group.

For attempts per question, we have five significant effects, starting with a main effect for group (F(1,27) = 12.704, p=.001). Figure 3 makes clear that the CF group has more attempts per question than the NOCF group in all micro sessions. This seems to be an outcome related to the differences in conditions. We see at the same time a group by session effect (F(1,27) = 12.065, p = .002) in combination with a session effect F(1,27) = 8.165, p = .008), with a smaller difference between the two groups in session 2. A separate analysis of the two groups returns no effect for session in the NOCF group (F(1,12) = 1.061, p = .323) and a significant effect for the CF group (F(1,15) = 13.381, p = .002). This means that there is a change in attempt behaviour in the CF group but not in the NOCF group. The overall analysis gives an effect for both micro session (F(2,54) = 4.226, p = .020) and the interaction between micro session and session (F(2,54) = 4.049, p = .023). The differences between the two sessions in the second micro session seem to be the source of these effects. A post-hoc analysis (Sidak procedure) for session 1 reveals a significant difference between micro session S1-2 and the other two micro sessions, whereas no significant differences turn up at all in a post-hoc analysis of micro sessions in session 2.

The two group by session interaction effects above, for questions and attempts per question, suggest that the CF group, which had to formulate the correct answer before proceeding to the next question, became more proficient over time.

![Figure 3: Changes in practice behaviour. The first graph depicts mean video clips viewed for all participants (there is no group effect) for each 15-minute micro session. The second one depicts mean questions completed and the third mean attempts per question for each 15-minute micro session in each group.](image)

5. Discussion and Conclusions

Our analyses indicate that participants in both groups benefited from practice. Learners became increasingly confident about their ability to answer questions in the practice exercise and reported positive attitudes towards practice with the system throughout practice. However, their motivation levels tended to fall as they approached the end of each practice session. Given their positive attitudes and increasing confidence throughout the session, it seems plausible to attribute learners’ diminishing desire to practice fatigue. It is encouraging, however, to see that, although motivation decreased at the end of the session, their motivation returned to higher levels at the beginning of the second session. Post-questionnaire analysis point to a similar conclusion, as the results indicate that both groups had similarly positive views on individual questionnaire items surveying learners’ attitudes towards practice, opinions concerning utility of practice, and perceived difficulty of the videos and questions. Analysis of the practice logs suggests that participants in the CF group had to produce many more utterances, but that this requirement did not cause significant differences in their
evaluation of the system or their motivation during practice.

Our findings here suggest that CF in our particular computer environment, in the context of grammar practice, did not have a significant effect on learner motivation. This stands in contrast to some views in the SLA literature which suggest it would have a negative effect (e.g. [13]). It may be the case that learning context has a role to play. In cases where learners communicate in meaningful exchanges frequent CF may be perceived as disruptive to the natural flow of exchange. In some social situations, such as the language classroom, CF may also cause a learner to feel embarrassed or to lose face and negatively impact learner motivation. In oral grammar exercises with a computer, which lacks a meaningful exchange or social element, CF may not have the same negative effect.

Other possible explanations include that two 45-minute sessions of practice may not have been enough time to see a group or group-time effect emerge, or that our NOCF group condition may not have been as poor as we assumed: This might be in part due to us telling participants that their recorded answers would be scored and that these scores would be provided to them at a later time. The fact that someone would listen to their utterances and score them, even if not immediately, may have been enough to motivate learners.

A final possibility is that the results were influenced by the experimental design. In the first session, learners completed a number of activities before commencing practice. For a non-native Dutch learner, the experience of filling out a background questionnaire, reading through instruction materials, and completing two proficiency tests before beginning practice may have a large and similar effect on both the CF and NOCF groups which hides any smaller CF-related effects. An interesting future possibility would be to structure the experiment so that questionnaires and proficiency tests are completed on separate days, with two days consisting only of practice activities, so that learners begin in more similar motivational states at S1-0 and S2-0.

Taken together, the results are encouraging. Participants in both groups maintained positive attitudes towards our system over time, and seemed to become increasingly confident about their ability to do well in the practice exercises. Their views after practice confirm this conclusion. Based on our results here, we believe this setup is suitable for situated motivation, by means of mini-questionnaires and practice log analysis, and that these data can be combined with pre- and post-test data to provide a compelling account of what occurs during practice, with links to outcomes. An interesting topic for future experiments would be to use our system to investigate different motivational strategies in the system to explore how learners’ motivation can be sustained over time.

6. Acknowledgements

We thank our colleague Joost van Doremalen for developing the ASR component of the CALL system used in this experiment, and the three anonymous reviewers for their comments. This work is part of the research program ‘Feedback and the acquisition of syntax in oral proficiency’ (FASOP), which is funded by the Netherlands Organisation for Scientific Research (NWO).

7. References