L2 syntax acquisition: the effect of oral and written computer assisted practice

Polina Drozdova, Catia Cucchiarini, Helmer Strik

Centre for Language and Speech Technology, Radboud University, Nijmegen, the Netherlands

p.drozdova@let.ru.nl, c.cucchiarini@let.ru.nl, W.Strik@let.ru.nl

Abstract
The purpose of the present study was to investigate the effect of modality-specific practice through a CALL system on the acquisition of an L2 syntactic phenomenon: subject-verb inversion. Two groups of learners of Dutch participated in the study: one group worked with a version of the CALL system that allows drag-and-drop exercises while another one practiced with a different version of the system that makes use of Automatic Speech Recognition (ASR) to provide practice and feedback on spoken utterances. Progress was measured by comparing the results of the groups on pre- and post-tests with discourse completion tasks in written and oral forms. The CALL system performed well, and statistical analyses revealed a significant difference between the results of all participants in pre- and post-tests, demonstrating the positive effect of practice with both versions of the program on the acquisition of the syntactic feature addressed in the training. Larger improvements were observed for oral training, but no significant effect of practice modality on the progress of the participants or their appreciation of the program was found. The implications of these findings are discussed.

Index Terms: practice modality, L2 syntax, CALL

1. Introduction
Standard language classes give little opportunity to students to practice actual spoken interaction. Learners are usually required to present an output of their activities in written form: from worksheets, where they have to write separate words and sentences to essays [1]. It remains unclear, however, whether practice in the written modality contributes to the development of speaking, since writing and speaking differ in the demands they put on working memory and the involvement of speech-motor mechanism [2, 3]. According to deKeyser [4], the practice effect is modality specific and there is little possible transfer even between tasks that appear to be similar at first glance. Following this line of reasoning and considering that writing and speaking require different cognitive processes, it is assumed that in order to improve speaking skills, oral training is necessary, whereas training in written form is needed to enhance writing skills. Studies on the effect of computer mediated communication (CMC) on the development of oral proficiency [3, 5, 6, 7], on the other hand, demonstrated that chat interactions in written form can indirectly improve oral proficiency, thus suggesting that there is a transfer between writing and speaking and that practice in one modality can indirectly improve performance in the other.

In the present study we investigated the effect of modality of practice on the acquisition of a syntactic construction and subsequent performance in the same or different modality. To study this question we used an ASR-based CALL system for learning subject-verb inversion in Dutch with corrective feedback developed for the FASOP (Feedback and the Acquisition of Syntax in Oral Proficiency) project [8, 9], which was adjusted for the present research. The use of an ASR-based CALL application gave us the opportunity to present and obtain information in different modalities, to access the logs of the participants during experimental sessions, and to reduce learner anxiety which many L2 learners experience when they have to speak in a foreign language [10].

In Section 2 we describe the methodology adopted in this study conducted among learners of Dutch in June 2012. We start with the description of the methods used in the research and then discuss the results, focusing on the role of practice in general, influence of the modality of practice and learners’ impressions from working with a CALL system.

2. Method
The study was conducted among learners of Dutch studying to achieve the A2 or B1 level according to the Common European Framework of Reference (CEFR). In this section we present the experimental setting, participants and materials used in the study.

2.1. Participants and experimental procedure
Eighteen students (six males and twelve females) attending Dutch courses at the Language Centre of the University of Tilburg, ROC Tilburg and Radboud in’to Languages (the Language Centre of the University of Nijmegen) took part in the experiment. They received 10 Euro payment for their participation. The age of the participants ranged from 22 to 44. All participants had a university degree or were in the process of studying for a university degree, had been studying Dutch for at least 4 months and had been living in the Netherlands for more than half a year. The participants have diverse ethnic backgrounds and come from different countries. Most of them spoke foreign languages other than Dutch, and mentioned that they did not have many opportunities to use Dutch in daily life: in general they spent no more than one hour a day speaking, reading, writing or listening to Dutch outside the classroom.

The participants were randomly assigned to two experimental groups. The groups differed in the modality of practice: oral vs. written modality. Participants from the oral modality group had to pronounce the sentences to an ASR system, whereas in the latter case they had to perform drag and drop tasks during the training. They worked with a program individually in a separate study room in two experimental sessions, with each session lasting for about 90 minutes. Both sessions were done on different days. The first session included a pre-questionnaire, a pre-test and the first part of the training, a post-test and a post-questionnaire. Although the study focused on verb second in Dutch, which is the movement of the verb to the second position after a topicalized element, the rule was not explicitly
introduced to students and both tests and training contained fillers as well as target sentences.

2.2. Subject-verb inversion training

In the training session participants watched a series of videos developed by an educational publisher for L2 learners of Dutch, followed by questions about the plot of the story. The first experimental session included 14 video clips, followed by 58 questions. The second session consisted of 16 videos and 64 questions. Each question set incorporated 20 target questions, answers to which should include subject-verb inversion. To prime subject-verb inversion the first block of the target reply was given. One group of participants had to assemble the blocks with the mouse in a drag-and-drop exercise, whereas the other had to pronounce the answer to the system. Examples of the drag-and-drop and oral exercises are given in Figures 1 and 2.

![Figure 1](image1.png)

Figure 1: A screenshot of the drag-and-drop exercise of the written modality group. The learners had to answer the question: “Why does Melvin think that Tom is not fine?” by putting the word blocks in the right order.

![Figure 2](image2.png)

Figure 2: A screenshot of the training of the oral modality group. The learners have to record their answer to the question: “What is Melvin going to do?”

The program provided feedback to the user: red if the answer was incorrect, in this case the user had to try once again and after a second failed attempt he/she received one more block as a prompt; green if the word order was correct (in this case the participant could proceed to the next question), or white feedback when the program could not recognize the utterance in the oral modality, or the participant did not use all the necessary blocks in the drag-and-drop task. The example of the negative feedback is given in Figure 3.

![Figure 3](image3.png)

Figure 3: A screenshot of the red feedback: “That is not correct. Try once again”.

2.3. Pre- and post-tests

During pre- and post-tests the production accuracy of the participants was measured through a discourse completion test (DCT) in which they had to produce a complete sentence using the prompts from the program in the form of words and pictures. The participants were given the beginning of the sentence to ensure that the target construction would be used. To compare performance in the two modalities, the learners were given two forms of the test: in one test they had to pronounce the sentence to the program, and in the other they were required to type their response. Each test consisted of 32 questions, half of which were target sentences where subject-verb inversion was required, and half of which were fillers representing different problematic areas for learners in the target group, such as: word order in subordinate clauses, subject-verb agreement, use of verb phrases, presence of subject and verb, etc. Both tests had a time bar. Since the participants had restricted time for their answers and had to start speaking immediately, we hoped to increase spontaneity and intuitiveness of their responses. Since the experiment was conducted in two sessions, all the participants received four versions of the tests (two in the written modality and two in the oral modality), which were randomly distributed across the sessions. The reliability statistics conducted on the items from written and oral DCT revealed a high level of internal consistency of the tests: Cronbach’s α for the oral DCTs constituted .913, and for the written DCT .918.

The tasks for the oral DCT were piloted in previous experiments using the same program [8]. The tasks for the written DCT were additionally developed and were piloted prior to the main study.

2.4. Users’ questionnaires

The pre-questionnaire consisted of fifteen multiple choice and short open-ended questions, including personal questions about the participants’ age, gender, mother tongue, foreign languages educational level, country of origin, Dutch L2 learning. The pre-questionnaire was the same for both groups.
The post-questionnaire was developed to investigate the participants’ impression about using the program. Since the two groups used different versions of the CALL system, they received different versions of the post-questionnaire. The questions were in the form of Likert statements to which participants replied on a five-point scale indicating whether they (1) strongly disagree, (2) disagree, (3) neither agree or disagree, (4) agree, or (5) strongly agree with the statement. The questions reflected the general impression about working with the program, activity, materials and feedback. Finally, the participants were asked to make suggestions about how the program could be improved and give overall comments.

2.5. Data analyses

The progress of the participants was recorded in program logs. The target utterances produced in the oral DCT were transcribed and assigned 1 or 0, depending on the participant’s answer, and the same procedure was used in evaluating the data from the written DCTs. Scores of the participants were added together and divided by the total number of utterances. The obtained numbers were used in data analyses. Descriptive statistics were computed to report the average results on the pre-tests and post-tests for the oral and written DCTs for the whole sample and different groups. In order to evaluate the improvement of the scores and its correlation with the modality of the training, a repeated-measures ANOVA was conducted.

The answers to the post-questionnaire were divided into two groups: 1) overall impression about the programme and 2) the feedback. Mean scores were calculated.

3. Results

This section presents the results of the quantitative analyses of the pre- and post-tests and the pre- and post-questionnaires.

3.1. The role of practice in the acquisition of L2 syntax.

The mean scores of the participants on the pre- and post-written and oral DCTs are displayed in Table 1.

Table 1. Participants’ scores on the oral and written DCTs

<table>
<thead>
<tr>
<th>Test type</th>
<th>Pre-test</th>
<th>Post-test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>min</td>
<td>max</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>written DCT</td>
<td>0.125</td>
<td>1</td>
</tr>
<tr>
<td>oral DCT</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

As can be seen from Table 1, the participants improved their results on both types of tests. The analysis revealed a significant difference between the mean scores of the participants in the pre-tests and the post-tests (F (1, 15) = 8.103, p < .05).

3.2. Modality of practice and L2 syntax acquisition

The groups participating in the study differed in the modality of practice they received with one group answering the questions assembling the word-blocks provided by the program with the mouse and the second group pronouncing the sentence to the program using the same word-blocks. The gains of the participants on different test types are reported in Table 2.

Table 2. Gains in the written and oral DCT for the oral and written training groups

<table>
<thead>
<tr>
<th>Test type</th>
<th>oral training</th>
<th>written training</th>
</tr>
</thead>
<tbody>
<tr>
<td>written DCT</td>
<td>0.09 (SD = .14)</td>
<td>0.04 (SD = .16)</td>
</tr>
<tr>
<td>oral DCT</td>
<td>0.11 (SD = .24)</td>
<td>0.09 (SD = .12)</td>
</tr>
</tbody>
</table>

It appears from the table that both groups showed more improvement on the oral DCT than on the written DCT, and the mean gain scores were higher for the oral training group on both oral and written DCT.

To further study the effect of the type of task on the development of syntactic knowledge we compared the gains of the participants in written and oral discourse completion tasks, using a repeated-measures ANOVA with time (pre-post-tests) and test type (written DCT-oral DCT) as within-subject factors and modality (written-oral) as a between-subject factor. Levene’s test has demonstrated that variance were homogeneous for all levels of repeated measures variables.

Statistical analyses revealed no main effect of test type (F (1, 15) = 2.166, ns). There was no significant interaction effect between type and time of test (F (1, 15) = 0.707, ns). No main effect of the modality of training was observed (F (1, 15) = 0.676, ns). There were no significant two-way interactions between the time of the test and the modality of the training (F (1, 15) = 0.079, ns). or between the type of the test and the modality of the training (F (1, 15) = 0.252, ns). The three-way interaction between all three variables: time of the training, type of task and the modality (written-oral) was not significant either (F (1, 15) = 0.015).

3.3. The appreciation of the CALL application

In general all participants were positive about the program (M=3.61, SD = .47) and about feedback provided by the system (M=3.45, SD = .58). One-way ANOVA revealed no significant effect of the modality of training on the overall appreciation of the program (F (1,15) = 0.261, ns) or the feedback (F (1,15) = 0.05, ns).

The qualitative analyses of the responses to the open ended questions revealed that for the oral DCT group negative comments were more often associated with the feedback, whereas the written DCT group would have appreciated to have more opportunities to speak. The observations are in line with the tendencies observed in the data. Even though the effect of the modality of training was not significant on these statements, participants who received the training in the oral modality agreed more often with the statement: “Sometimes the system said that I made a mistake when it was not the case” (M = 3.22, SD = 1.2) than the written modality group.
(M = 2.88, SD = 1.46). At the same time, even though the difference was not statistically significant, more people from the written training group indicated that they would prefer to speak their answers rather than to type them (M = 3.38, SD = 1.41), in comparison to the oral training group (M = 2.44, SD = 1.01).

4. Discussion

In the present study we investigated whether and how oral and written practice through a CALL system contributes to the acquisition of a specific L2 syntactic construction and whether the effect is modality specific or generalizes across different modalities.

We observed that both forms of training with the CALL system were beneficial for the acquisition of subject-verb inversion in Dutch L2 since the scores of the participants on the post-test were significantly higher than their scores on the pre-test. This supports the hypothesis that productive practice with meaningful, comprehensive input and immediate feedback improves performance. These results also support our observations that both versions of the CALL system worked well, and were suitable for such productive practice with immediate feedback.

However, no significant effect of modality of training was observed on L2 acquisition. This points at the possibility of a cross-modality transfer, as demonstrated in studies comparing the effects of computer mediated communication and oral classroom interaction on the development of oral skills [3, 6, 11, 12].

Although both groups improved on oral and written tasks, irrespective of the modality of practice, and no interaction was found between the time and type of the test and the modality of the training, descriptive statistics indicate a tendency of the oral training group to outperform the group that received drag-and-drop training. The lack of significant differences may be ascribed to certain limitations of the study such as small sample size, limited length of the training and relatively high performance of the participants on the pre-tests. More data, obtained with tasks of more variable difficulty, would be needed to determine whether these results are suggestive of some tendencies, such as the advantage of oral modality of training over written one, the transfer of skills across modalities, or the larger improvement on modality consistent tasks.

Based on studies of learners’ beliefs demonstrating that second language learners consider that development of speaking proficiency should be the focus of language instruction [13], and assuming that performing communicative tasks with the computer making use of ASR should be a new activity type that can reduce learner anxiety, we hypothesized that the participants who received oral training would be more positive about the program than their peers who were assigned to the drag-and-drop group.

On the other hand, even though the performance of the speech recognizer had been optimized, the participants in the oral training group could still be confronted with erroneous feedback (false acceptances when an incorrect utterance was classified as correct, and false rejections when the correct utterance was classified as incorrect [14]) whereas for the drag-and-drop group feedback was always accurate.

These assumptions were only indirectly supported by the data. The appreciation of the program or the feedback did not differ significantly between the groups, but the group trained with drag-and-drop exercises mentioned the necessity of oral tasks in the training session, whereas the oral modality group was more critical about the feedback in the open-ended questions in the post-questionnaires.

This result supports the observation made by Yang [13], whose study indicated that for the participants it was important to address all four language skills: speaking, reading, listening and writing. Working with the system included practicing all four skills for both groups: writing in questionnaires and written DCTs, speaking in oral DCT and training session for the oral training group, listening in the training session and reading as participants had to read the tasks, tutorials and prompts. However, only the group that received the training with drop and felt the need for more oral practice as the task they had to accomplish was less communication-oriented.

Finally, the use of computer software for research of this type was shown to be advantageous both for the learners as they can practice on their own without feeling anxious speaking in a foreign language, and for the researcher as all the output produced by the students is easily accessible.

5. Conclusions

In the present study we investigated the impact of different types of training with a CALL system on the acquisition of an L2 syntactic construction, and participants’ appreciation of the program. The results demonstrated positive effects of communicative practice on the improvement of syntactic proficiency in the learners and the generalization of the effect across modalities, since no significant effect of the modality of the training on the participants’ performance on the different types of tasks was revealed.

Participants from both groups were, in general, positive about the training with the program, which points at the advantages of using a CALL system as a medium of L2 practice, on the conditions that the feedback provided by the program is accurate in most of the cases, and training provided is variable and oriented towards the development of all four language skills: reading, writing, speaking and listening.

Further research with a larger sample of participants, with lower proficiency, and therefore lower performance at the pre-test, different focus of training, and tasks with more variable complexity is necessary to investigate the tendencies observed in this study.

6. Acknowledgements

We would like to thank our colleague Joost van Doremalen for developing the ASR-based CALL system used in this experiment, Roeland van Hout for advice on statistics, Bart Penning de Vries and Steve Bodnar for allowing us to use the materials they developed for the research project ‘Feedback and the acquisition of syntax in oral proficiency’ (FASOP), which is funded by the Netherlands Organisation for Scientific Research (NWO).

7. References


