

# Spoken Output and Corrective Feedback in an ASR-based CALL system

B.Penning de Vries, S.Bodnar, C.Cucchiarini, R.van Hout, H.Strik

Centre for Language and Speech Technology, Radboud University Nijmegen

## *Introduction*

Speech technology offers exciting possibilities for language learning research. Through implementation of automatic speech recognition (ASR) in Computer Assisted Language Learning (CALL) systems, second language (L2) learners can be provided with speaking practice that is individual, causes less learner anxiety than in face to face communication, without time limits, and with the possibility to adapt to learner differences - options that are not available to the learner in the classroom (cf. Truscott 1999). Golonka et al. (2012) find that systems offering spoken practice and feedback are effective for pronunciation, but a recent review showed that there no systems that offer grammar practice in the spoken modality (Bodnar et al. 2011). For second language acquisition (SLA), research on the acquisition of spoken grammar is challenging. CALL systems that make use of ASR to enable spoken practice are well suited to research this issue (Penning de Vries et al. 2010). Because learners interact with a computer system individually, much of the learning context can be controlled. As such, the qualities of CALL allow for detailed inspections of effects of input, output, and task design on proficiency. In addition, CALL can log learner behavior during the task. In this paper we present a language learning experiment using our ASR-based CALL system.

## *Research background*

Second language acquisition (SLA) requires language input, but the L2 is often not fully acquired as some features in the input may not be salient to the learner. Important in this respect is the notion of ‘noticing’ (Schmidt 1990), which posits as a requirement for acquisition that an L2 feature is noticed. Noticing can be achieved through providing corrective feedback (CF), but also through production of output. The role of output in language learning is that it provides learners with skill-specific practice, opportunities to notice gaps in the interlanguage, and may prompt them to reflect on a metalinguistic level (De Bot 1996, Swain 1985, DeKeyser 2007). The role of CF has been heavily researched, with several meta-analyses (Norris Ortega 2000 through to Lyster & Saito 2010) pointing towards a beneficial effect of CF. However, Russell and Spada (2006) conclude that more research is needed on what factors mediate in the effectiveness of CF. Studies report different effects depending on the type of CF, individual differences, type of target structure, and research context (Lyster, Saito & Sato 2013). It is here where a methodical approach through an ASR-based CALL system can provide insight into how these factors interact (cf. Goo & Mackey 2013). Our CALL system provides learners with a spoken exercise on Dutch word order, and it is embedded in a language experiment to evaluate its effectiveness. In this way we address outstanding issues in SLA, namely the effect of output practice and automatic corrective feedback (CF) on spoken language proficiency. Our research question asks: is there a difference in effectiveness between output practice with automatic CF and output practice that relies on self-monitoring (i.e. without CF)?

## Method

The main part of the experiment is the speaking exercise (see screenshot Figure 1) on Dutch word order. The target feature is Dutch V2, a difficult feature for L2 speakers of Dutch. In this exercise, the learners watch a film clip of 35 seconds, and are then asked questions about it. They answer by using given 'word blocks' to construct a sentence. This is necessary for high ASR accuracy to provide pedagogically sound CF. Analyzing non-native input is challenging for the ASR, but a pilot experiment showed that feedback accuracy was high (96%). The system reacts to the learners' utterances only in the CF condition: in the other condition, it only logs the utterance. As such, the two experimental conditions are 1) a group receiving automatic CF on their utterances (CF group), and 2) a group with only speaking practice (NO-CF group). The groups performed two pre- and post-tests for a complementary picture on the participants' proficiency (cf. Norris Ortega 2003): a grammaticality judgment task for receptive knowledge (GJT), and a discourse completion task (DCT) to measure spoken production accuracy (cf. Ellis 2005). During the two training sessions of 45 minutes, the system logs the user-system interactions, which allows insight into the sessions.



Figure 1. Screenshot of the speaking practice activity, GREET.

## Results

In our experiment, 31 language learners with 15 different L1s practiced individually with our system. In the post-questionnaire, both groups evaluated the system positively on a five-point Likert scale (CF mean=4.14, SD=.40; NOCF mean=3.74, SD=.69), indicating that they found the practice enjoyable, and useful. In the proficiency tests we found that the GREET system is effective in improving the participants' knowledge of V2 in both conditions. The learners improve significantly on the target structure, in both proficiency tests. For the filler items, there is no increase (see Figure 2), showing that increase is not due to task familiarity.

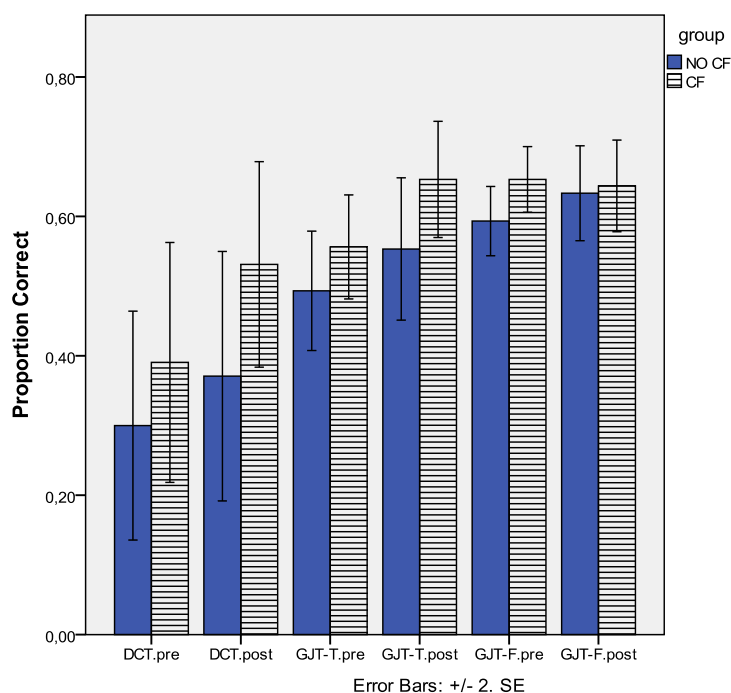


Figure 2. Bar graph of the pre- and post-test in our system, with 2 standard error bars. It shows the discourse completion task (DCT), and the grammaticality judgment test targets (GJT-T) and fillers (GJT-F).

However, we do not find a significant interaction of group by treatment, which suggests that output practice with or without CF did not have a different effect on proficiency. We turned to a more detailed analysis to answer our research question, and examine the logs of the treatment sessions. We found significant differences between the groups with respect to the number of attempts they practiced per question. We looked at the proficiency tests and how they correlate with learner behavior in the two practice conditions. Additionally we discuss the individual differences (such as proficiency level and computer literacy) that are found in our participant group. In this way, our paper presents a controlled study on the role of output and CF, and discusses factors that are found to influence the effectiveness spoken language practice. We look forward to discussing these results at the conference.

### References

- Bodnar, S., Cucchiarini, C. & Strik, H. (2011). Computer-assisted grammar practice for oral communication. Proceedings of the 3rd International Conference on Computer Supported Education, Noordwijkerhout, The Netherlands, 355-361.
- De Bot, K. (1996). The psycholinguistics of the Output Hypothesis. *Language Learning*, 46, 529-555. doi: 10.1111/j.1467-1770.1996.tb01246.x
- DeKeyser, R. (2007). Introduction: Situating the concept of practice. In R. DeKeyser (Ed.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology*, (pp. 1-19). Cambridge, USA: Cambridge University Press.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27, 141-172. doi:10.1017/S0272263105050096

- Golonka, E., Bowles, A., Frank, V., Richardson, L. & Freynik, S. (2012). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer Assisted Language Learning*, 1-36. doi:10.1080/09588221.2012.700315.
- Goo, J., & Mackey, A. (2013). The case against the case against recasts. *Studies in Second Language Acquisition*, 35, pp 127-165. doi:10.1017/S0272263112000708.
- Lyster, R., Saito, K. (2010). Oral feedback in classroom SLA: A meta-analysis. *Studies in Second Language Acquisition*, 32, 265–302. doi:10.1017/S0272263109990520.
- Lyster, R., Saito, K., Sato, M. (2013). Oral corrective feedback in second language classrooms. *Language Teaching* 46, 1-40. doi:10.1017/S0261444812000365.
- Norris, J. & Ortega, L. (2003). Defining and measuring SLA. In C. Doughty & M. Long (Eds.) *The Handbook of Second Language Acquisition* (pp. 717-761). Malden: Blackwell.
- Norris, J., Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, 50, 417-528. doi: 10.1111/0023-8333.00136.
- Penning de Vries, B., Cucchiarini, C., van Hout, R., Strik, H. (2011) Adaptive Corrective Feedback in Second Language Learning. In S.De Wannemacker G. Clarebout, P. De Causmaecker (Eds.) *Interdisciplinary Approaches to Adaptive Learning. A Look at the Neighbours* (pp 1-14). Springer Verlag. doi: 10.1007/978-3-642-20074-8\_1
- Russell, J., & Spada, N. (2006). The effectiveness of corrective feedback for acquisition of L2 grammar: A meta-analysis of the research. In J. Norris & L. Ortega (Eds.) *Synthesizing research on language learning and teaching*. Amsterdam: John Benjamins. (pp. 133-164).
- Schmidt, R. (1990). The Role of Consciousness in Second Language Learning. *Applied Linguistics*, 11, 129-158. doi:10.1093/applin/11.2.129
- Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in Second Language Acquisition* (pp. 235-253). Rowley, MA: Newbury House.
- Truscott, J. (1999). What's wrong with oral grammar correction. *Canadian Modern Language Review* 55, 437-56. Retrieved from [http://www.hss.nthu.edu.tw/~fl/faculty/John/What's Wrong with Oral Grammar Correction 1999.htm](http://www.hss.nthu.edu.tw/~fl/faculty/John/What's%20Wrong%20with%20Oral%20Grammar%20Correction%201999.htm)