Production training benefits both comprehension and production of grammatical gender agreement in L2 German

Valérie Keppenne (The Pennsylvania State University), Elise W. M. Hopman (University of Wisconsin-Madison), & Carrie N. Jackson (The Pennsylvania State University)
vxk57@psu.edu

Recent memory research suggests that, at least at the single word level, language production provides a stronger learning experience than language comprehension due to the different memory processes involved (Karpicke & Roediger, 2008). Taking these findings, Hopman and MacDonald (2018) investigated the benefits of production training in an artificial language with a manipulation of learning condition (comprehension vs. production) followed by comprehension tests. The results showed that participants with production-based training were more accurate and faster on post-training comprehension tasks targeting semantic and number agreement features than participants with comprehension-based training. However, the generalization of production skills to the domain of comprehension in natural L2 learning remains contested (DeKeyser, 2007; VanPatten, 2013). Their study also did not test participants' production abilities, despite possible differences compared to comprehension abilities. Furthermore, results from lab-based artificial language studies do not always replicate among actual beginning classroom-based learners (Paul & Grüter, 2016), especially because the grammatical feature paradigm of natural languages is often more complex than that of artificial languages.

To test whether findings from Hopman and MacDonald (2018) extend to classroom-based learners, the present study adapts their training paradigm to test the learning of grammatical gender among beginning L2 learners. First-semester American L2 German learners received production- (*N*=21) or comprehension-based (*N*=20) training on grammatical gender assignment and agreement between articles, adjectives, and nouns (*ein*$_{MASC}$ *blauer*$_{MASC}$ *gepunkteter*$_{MASC}$ *Becher* 'a blue dotted cup') for 15 non-cognate German nouns via computer-based training. Both groups received equal amounts of auditory and written input with pictures, with training blocks progressing from simple noun phrases to blocks with full sentences. Each block of passive exposure was followed by a block of active learning: The comprehension group completed a sentence-picture matching task; the production group completed an oral picture-description task (Figure 1). Immediately after training, both groups completed a forced-choice (FC) gender suffix comprehension task, an error monitoring task (EM), and a written picture-description task, all targeting articles and gender agreement markers (Figure 2).

Post-training comprehension measures showed no significant reaction time differences between groups. However, significant differences emerged with regard to comprehension accuracy (Figure 3). For the FC task without nouns, the production group was more accurate than the comprehension group (*M*=82% vs. 69%, *p*=.002). For the EM task, the production group was significantly more accurate than the comprehension group, as measured by d' scores (*M*=1.09 vs. 0.61, *p=.02*). Additional analyses of this task showed that for sentences with errors, performance differed according to error location: The production group was better at detecting mismatching articles than the comprehension group (*M*=70% vs. 60%, *p*=.001), and both groups had difficulties detecting mismatching gender suffixes on adjectives (*M*=29% vs. 26%, *p*=.24). The production test (Figure 3) revealed that although both groups struggled with gender assignment and agreement, the production group produced more matching article-adjective-noun strings than the comprehension group (*M*=36% vs. 17%, *p*<.001).

We conclude that production-based training is more beneficial than comprehension-based training for both the comprehension and production of grammatical gender and gender agreement among classroom-based L2 German learners. This largely replicates findings by Hopman and MacDonald (2018), and suggests that even in natural language learning, skills based on production training are transferable to the domain of comprehension despite the greater complexity of agreement paradigms in natural languages.

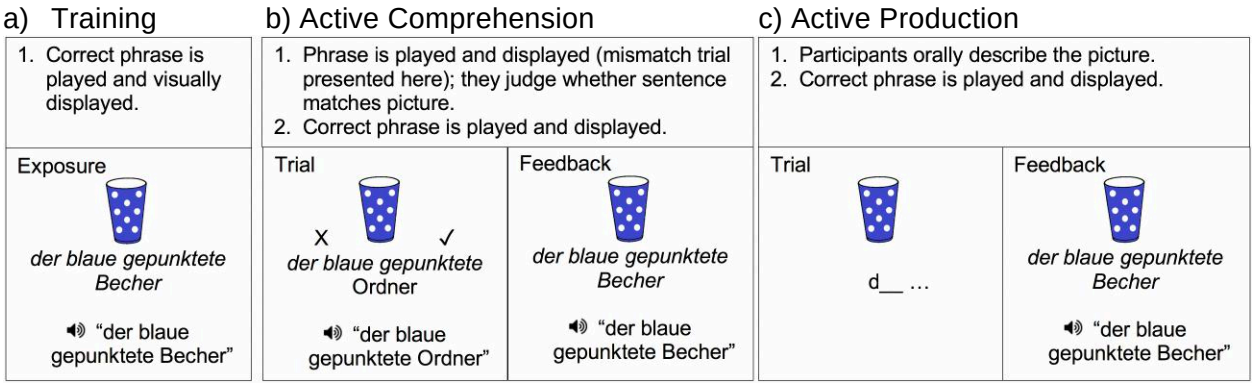

Figure 1. Sample items in each trial type from the training procedure.

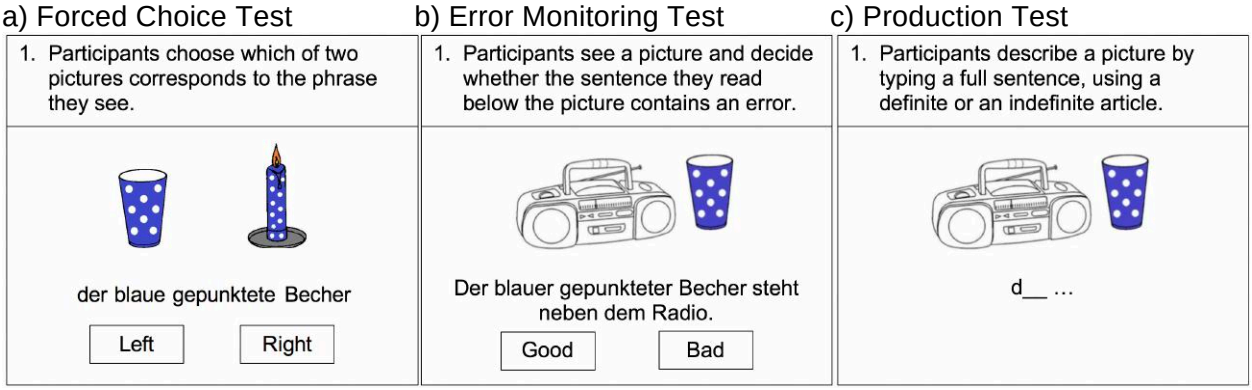

Figure 2. Sample items from the forced choice test, error monitoring test (mismatch trial), and production test.

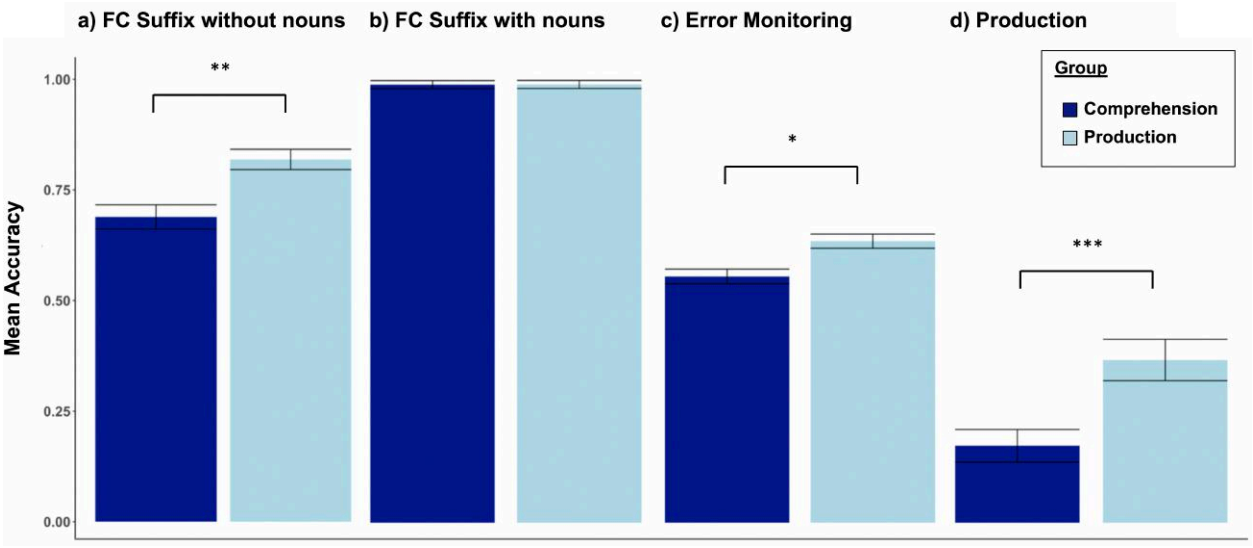

Figure 3. Mean accuracy across groups for comprehension and production. Error bars = 95% CI. C = Comprehension; P = Production. Significance levels: *p < .05, **p < .01, ***p < .001.