

Learning new vocabulary in German: the effects of inferring word meanings, type of feedback, and time of test

Shana K. Carpenter · Riebana E. Sachs · Beth Martin ·
Kristian Schmidt · Ruxandra Looft

Published online: 2 December 2011
© Psychonomic Society, Inc. 2011

Abstract In the present study, introductory-level German students read a simplified story and learned the meanings of new German words by reading English translations in marginal glosses versus trying to infer (i.e., guess) their translations. Students who inferred translations were given feedback in English or in German, or no feedback at all. Although immediate retention of new vocabulary was better for students who used marginal glosses, students who inferred word meanings and then received English feedback forgot fewer translations over time. Plausible but inaccurate inferences (i.e., those that made sense in the context) were more likely to be corrected by students who received English feedback as compared with German feedback, providing support for the beneficial effects of mediating information. Implausible inaccurate inferences, however, were more likely to be corrected on the delayed vocabulary test by students who received German feedback as compared with English feedback, possibly because of the additional contextual support provided by German feedback.

Keywords Human memory and learning · Language/memory interactions

Introduction

An important question in foreign language learning is how to optimize the retention of new vocabulary. It has been demonstrated that vocabulary can be acquired through reading (e.g., Saragi, Nation, & Meister, 1978). However, it is not clear exactly what readers should do when they encounter an unknown word. Should they directly access the word's meaning (e.g., by using a dictionary or marginal gloss)? Or, should they try to infer (i.e., guess) its meaning from the context?

One concern about trying to infer word meanings from context is that learners are not always going to infer correctly (e.g., Laufer & Sim, 1985; Nation, 2001). If learners infer the wrong meaning of a word, how does this error affect learning? The prominent view espoused by influential learning theorists is that self-generated errors might become reinforced and interfere with memory for the correct answer (e.g., Guthrie, 1942; Skinner, 1958). According to this view, learning would be best under conditions in which errors are avoided and students are exposed to only the correct information.

Whether or not this is true is difficult to glean from past research, however. Although some studies have shown that learners can successfully infer and retain the meanings of new words from reading (e.g., Horst, Cobb, & Meara, 1998; Nation, 2001; Pitts, White, & Krashen, 1989), other studies have shown that vocabulary gains are much greater when learners use dictionaries while reading (e.g., Grace, 1998; Knight, 1994; Luppescu & Day, 1993). In these studies, however, vocabulary was acquired *incidentally* while participants read the material to prepare for a comprehension test, without knowledge that vocabulary would be tested. When the goal is comprehension, participants may not deliberately try to

S. K. Carpenter · R. E. Sachs · B. Martin · K. Schmidt · R. Looft
Iowa State University, Ames, Iowa,
Ames, IA, USA

S. K. Carpenter (✉)
Department of Psychology, Iowa State University,
W112 Lagomarcino Hall,
Ames, IA 50011-3180, USA
e-mail: shacarp@iastate.edu

learn the meanings of unknown words (e.g., Hulstijn, Hollander, & Greidanus, 1996; Laufer, 2003). Indeed, some research has found that unknown words are more likely to be ignored when they are not relevant to comprehending the meaning of a passage (e.g., Hulstijn, 1993; Peters, Hulstijn, Sercu, & Lutjeharms, 2009).

Two studies used *intentional* learning paradigms to ensure that students deliberately tried to learn new foreign words by guessing their meanings. Forlano and Hoffman (1937) found that English-speaking students learned Hebrew words better if they were told the English translations of those words, as compared with first trying to guess the translations and then being given the translations. However, Parlow and Berlyne (1971) observed the opposite result. In their study, students learned the meanings of Turkish words better if they tried to guess their meanings and were then given the English translations, as compared with being given the translations without ever trying to guess. Although students in both studies almost always guessed the *wrong* English translation, one study found that these errors were harmful to learning, and the other did not.

A key difference between these two studies is that Parlow and Berlyne (1971) provided a contextual clue to guide students' guesses (e.g., *Hafiz* is a type of animal), whereas Forlano and Hoffman (1937) did not. This clue could have increased the chances that students' incorrect guesses would be more plausible and belong to the same category as the correct answer. Indeed, recent research on human memory has shown that students are more likely to correct self-generated errors that are *related* to, rather than unrelated to, the correct answer (e.g., Huelser & Metcalfe, [in press](#)). It is possible, therefore, that when students infer the wrong meaning of a word in a foreign language, the plausibility of that error could be an important determiner of whether students will correct it. This question has not been explored in any of the known research on this topic. It is also not clear how these errors might affect both short- and long-term learning, since previous studies on error correction have typically measured only immediate retention.

The purpose of the present study was to explore how much German vocabulary students can learn by trying to infer the meanings of new German words from context, and how often wrong inferences are corrected after both immediate and delayed time intervals as a function of: (a) the plausibility of the error, and (b) the type of feedback that is provided to correct it. Introductory German students read a story and learned the meanings of new German words by either trying to infer the English translations from the story, or by reading the English translations in marginal glosses. Students who inferred translations received feedback in one of the following ways: (a) English translations, (b) simplified descriptions of the words in

German, or (c) no feedback at all. All students were then given a vocabulary test (requiring provision of the English translation of each German word) over half of the words immediately, and the other half after a 4-day delay.

Method

Participants

Eighty undergraduate students from introductory German courses at Iowa State University participated for extra credit. Twenty students were each randomly assigned to learn new words through (a) inferring + English feedback, (b) inferring + German feedback, (c) inferring + no feedback, or (d) marginal gloss.

Materials, design, and procedure

All students read a simplified version of a German children's book (Lindgren, 1980). The story contained approximately 600 words, including 16 target words that were judged by the instructor to be unfamiliar to students prior to the experiment (e.g., *der Verkehr*: "traffic"). The materials are available from the authors upon request.

During their regularly scheduled class time, each student received a packet containing a consent form and general instructions on the first page. These instructions informed participants that they would be reading a short story in German that was intended to teach them the meanings of some new German words, over which they would later be tested. The second page of the packet contained a pretest over the 16 target words. After completing the pretest (which required students to write down the English translation for each German word), students began reading the story.

For students in the infer conditions, the target words were not marked or identified in any way. Students were simply asked to read the story once to get acquainted with it. Students in the marginal gloss condition were given the same general instructions, but for these students, the 16 target words were underlined within the story, and an English translation was provided in the margin next to each target word.

All students then read a second copy of the story, this time with the target words underlined. For students in the infer conditions, the target words were listed in the bottom margins of the pages on which they appeared, and students were asked to write down a guess about the English translation of each word. For students in the marginal gloss condition, the target words were underlined with English translations in the margins, just like before. This time, however, the target words also appeared in the bottom

margins, and students were asked to copy down their English translations. This way, students in all conditions performed the same act of writing down English translations for the target words.

Immediately after reading the story twice, students in the infer + English feedback condition turned the page to see a list of the 16 target words with their English translations. For students in the infer + German feedback condition, this page listed the 16 target words with simplified descriptions in German (e.g., translated into English, *der Verkehr: Many cars on the street. Los Angeles has a lot of Verkehr*). Students in both conditions were informed that the information contained the correct meanings to the words that they had tried to infer, and that they should use this information to try to understand the meanings of the words and see if they inferred correctly.

The next page then contained a comprehension test over the story. For students in the marginal gloss and infer + no feedback conditions, the comprehension test followed immediately after the second reading of the story. For all students, the comprehension test consisted of five multiple-choice questions in German. These pertained to aspects of the story that were unrelated to the target words (e.g., *How old are the children? Where do they live?*).

The last page of each student’s packet then contained a vocabulary test over a group of eight target words that was uniquely randomly selected for each individual participant. Students were given the German words and were asked to write down their English translations. Students completed the experiment at their own pace and were monitored to ensure that they did not look back through the materials to answer the test questions. Altogether, the procedure lasted approximately 35 to 45 min.

Four days later, during their regularly scheduled class time, students were given a vocabulary test over the remaining eight target words. Again, students were given a page listing the German words and were asked to write down their English translations. At the bottom of the page was a question asking students to indicate which words, if any, they had thought about or looked up during the last 4 days. On an individual basis, any words that students identified as having thought about or looked up were eliminated from analyses. Upon completing this test, students were thanked and debriefed.

Results

Scoring

Responses on the pretest, inference, and vocabulary tests were awarded full credit (two points), partial credit (one point) or no credit (zero points). In addition to accuracy,

each inference was scored for plausibility—whether or not it made sense in the context—with one point awarded for plausible inferences, and zero points awarded for implausible inferences. To qualify as plausible, a wrong inference had to: (a) be the same word type as the correct answer, and (b) produce a semantically correct sentence when inserted in place of the correct answer. For example, consider the following section of the story (translated into English):

When they cross the street, Peter and Lena pay attention. “First you must look to the left. Then you must look to the right, and then again to the left,” Peter says to Lena. Peter pays attention to the Verkehr, because there are many cars driving on the street.

Examples of participants’ plausible but wrong inferences for *Verkehr* (“traffic”) were “stoplight,” “crossing guard,” and “drivers.” Examples of implausible wrong inferences were “to play” and “to shop” (verbs instead of nouns), or “leaves,” “building,” and “bird” (nouns that did not create a semantically correct sentence when inserted in place of *Verkehr*).

Two native speakers of German scored all of the responses for 25% of the participants in blind fashion. Inter-rater correlations were highly significant for accuracy as well as plausibility [all *rs* = .87 or higher, *ps* < .001], so the remainder of the responses were scored in blind fashion by one rater.

Pretest and initial inference

Performance on the pretest ranged from 1% to 2% across groups, indicating very little prior knowledge of the target words. One-way, between-subjects ANOVAs revealed no significant differences in pretest performance across groups, *F* = 1.97, and no significant differences in the accuracy and plausibility of initial inferences across the three infer groups, *F*s < 2.48 (Table 1).

Comprehension test

Participants in the marginal gloss group scored lower on the comprehension test than participants in the infer groups (Table 2). A one-way, between-subjects ANOVA revealed

Table 1 Mean proportion of words and their standard deviations inferred accurately and plausibly according to group

| | Group | | | | | |
|----------------------|------------------|-----------|-----------------|-----------|-------------|-----------|
| | English Feedback | | German Feedback | | No Feedback | |
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Proportion accurate | .23 | .10 | .26 | .11 | .20 | .08 |
| Proportion plausible | .57 | .13 | .54 | .13 | .57 | .13 |

Table 2 Mean proportion of comprehension test questions and their standard deviations answered correctly according to group

| Group | | Infer + English Feedback | | Infer + German Feedback | | Infer + No Feedback | |
|-------|-----|--------------------------|-----|-------------------------|-----|---------------------|-----|
| M | SD | M | SD | M | SD | M | SD |
| .74 | .22 | .82 | .23 | .91 | .10 | .86 | .15 |

that this effect was significant, $F(3, 76) = 3.10, p = .031, MSE = .033, \eta^2 = .109$. Reading a story containing English translations in the margins therefore appears to impair comprehension, possibly due to the act of dividing attention between the story and marginal glosses (e.g., see Cheng & Good, 2009).

Vocabulary tests

Figure 1 displays accuracy on the immediate and delayed vocabulary tests across the four groups. A 4 x 2 (group x time) mixed ANOVA revealed a significant main effect of group, $F(3, 76) = 35.91, p < .001, MSE = .037, \eta^2 = .586$. Tukey’s HSD test revealed that the marginal gloss and infer + English feedback groups significantly outperformed the other two, but did not differ from one another, and performance for the infer + no feedback group was significantly worse than that in the other three.

A main effect of time also emerged, $F(1, 76) = 98.47, p < .001, MSE = .023, \eta^2 = .564$, indicating that performance declined over the 4-day interval. Performance declined more sharply for participants in the marginal gloss and infer + English feedback groups as compared with the other two, $F(3, 76) = 24.91, p < .001, MSE = .023, \eta^2 = .496$, possibly due to the fact that immediate performance was lower for the latter two groups and less susceptible to further decline.

Although performance on the immediate test was highest for the marginal gloss group, this group’s performance dropped below that of the infer + English feedback group on the delayed test. A 2 x 2 (group x time) mixed ANOVA revealed that this interaction was significant, $F(1, 38) = 5.02, p = .031, MSE = .024, \eta^2 = .117$. The difference in performance between the immediate and delayed tests revealed that participants in the marginal gloss group forgot more translations over time ($M = .51, SD = .23$) than participants in the infer + English feedback group ($M = .35, SD = .20$), $t(38) = 2.24, p = .031, d = .708$.

Learning from errors

Out of all of the participants who wrote down an inference that was completely incorrect while reading the story, we

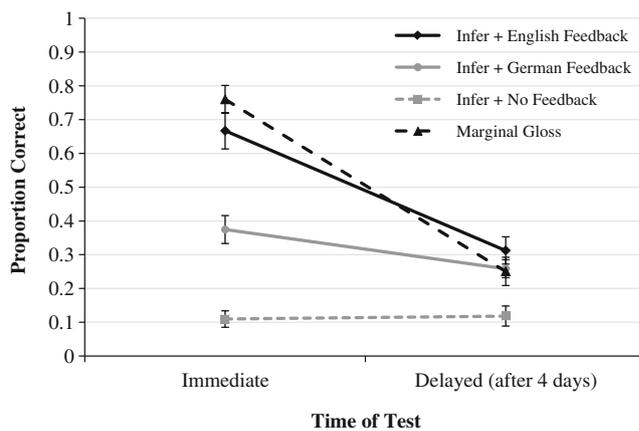


Fig. 1 Mean proportion of English translations recalled on the immediate and delayed vocabulary tests as a function of learning method. Error bars represent standard errors

calculated the proportion of participants who corrected one or more of these errors on the vocabulary tests as a function of the type of feedback that they received (English vs. German) and the plausibility of their initial wrong inference (plausible vs. implausible). These proportions are displayed in Table 3.

The proportion of participants who corrected plausible errors was greater for those who received English feedback than for those who received German feedback. A 2 x 2 chi square analysis revealed that this difference was significant at both the immediate ($\chi^2 = 5.23, p = .048$) and delayed ($\chi^2 = 6.06, p = .022$) vocabulary tests. The proportion of participants who corrected implausible errors on the immediate vocabulary test did not differ significantly between the two feedback groups ($\chi^2 < 1$). However, on the delayed test, the proportion of participants who corrected implausible errors was greater for those who received German feedback than for those who received English feedback, $\chi^2 = 4.25, p = .084$ (all p values reflect two-tailed tests).

Table 3 Proportion of participants who corrected at least one error as a function of the type of error, type of feedback, and timing of the vocabulary test

| Feedback | Type of Error | | | |
|----------|----------------|--------------|----------------|--------------|
| | Plausible | | Implausible | |
| | Immediate Test | Delayed Test | Immediate Test | Delayed Test |
| English | .80 | .74 | .82 | .29 |
| German | .45 | .33 | .68 | .65 |

Proportions are based on the number of participants who originally inferred the wrong meaning for at least one German word during reading. Given the fact that some participants did not produce errors of each type (e.g., some participants did not produce any implausible wrong inferences), the number of participants contributing to each of these proportions was not always all 20, but was never fewer than 17

Errors were seldom repeated by participants who received feedback. Across the various comparisons (e.g., plausible errors that were repeated on the immediate test), a maximum of only four participants who received feedback ever repeated an error. On the contrary, the majority of participants in the no feedback group repeated errors that they had made previously. On the immediate vocabulary test, 90% of participants in this group repeated at least one plausible error, and 72% repeated at least one implausible error. These proportions were reduced on the delayed test, but were still higher for plausible errors (26% of participants) than for the implausible errors (12% of participants). Only three participants in the no feedback group ever corrected any errors.

Discussion

These results help shed light on how students' word-learning strategies during reading affect vocabulary retention in a foreign language. When students tried to infer the meanings of unknown German words and then received English translations as feedback, they remembered those translations at a comparable rate overall to those who were exposed to the correct answers the entire time through marginal glosses. In fact, the former group retained more of the correct translations over time, despite the fact that they had less overall exposure to the correct answers and inferred the wrong meanings of the words at a rate of nearly 80% during reading. Thus, contrary to concerns that have been raised by learning theorists in the behaviorist tradition (e.g., Guthrie, 1942; Skinner, 1958), producing errors during learning does *not* appear to impair memory for the correct answer, as long as corrective feedback is provided (e.g., see also Kang, Pashler, Cepeda, Rohrer, Carpenter, & Mozer, 2011).

These results further demonstrate that the likelihood of correcting errors depends on the type of feedback that is provided, as well as on the plausibility of the error. When participants inferred a translation that was wrong but *plausible*, those who received English feedback were more likely to correct it than were participants who received German feedback. In addition to being the same word type, a plausible error and the correct translation are likely to share semantic features, making it difficult for German feedback to distinguish the error from the correct answer because it may contain some information that is relevant to both. For example, participants might plausibly but incorrectly infer that *der Verkehr* means “stoplight,” and German feedback (“Many cars on the street. Los Angeles has a lot of Verkehr”) may not provide enough information to lead participants from “stoplight” to “traffic” as the correct answer. English feedback, on the other hand, provides the correct answer unambiguously.

By sharing semantic features with the correct translation, a plausible error might act as mediating information that provides a link between the foreign word and its meaning. Such a “mediator effectiveness hypothesis” has been offered to explain the benefits of guessing the meanings of unknown words (e.g., Parlow & Berlyne, 1971), the tendency for participants to correct errors more often when they are related to the correct answer rather than when they are unrelated (e.g., Huelser & Metcalfe, *in press*), and the tendency for learning to benefit more from the act of retrieval than from restudying (e.g., Carpenter, 2011; Pyc & Rawson, 2010). Mediating information can enhance memory only if it is effectively linked with the correct answer, however. When feedback is provided in German, it may be difficult to link the mediator to the correct answer because it is less obvious what the correct answer is. Thus, the mediator effectiveness hypothesis could explain the tendency for participants to learn more from plausible errors that are followed by English feedback than by German feedback.

If an error is *implausible* and makes no sense in the context, this would indicate that the meaning of the word was not even remotely evident from the reading. So, any knowledge about this word's meaning is largely acquired from the feedback. German feedback provides a simplified description of the word, whereas English feedback provides only the translation. If feedback is the primary source of learning, then this creates a situation in which some participants learn the word meanings with the help of context (German feedback), and others do not (English feedback). Studying a list of translations out of context may encourage a rote learning strategy that can result in fairly good memory for those translations on an immediate test. However, learning with the help of context is known to produce advantages in long-term retention (e.g., Gipe, 1980). The extra contextual information provided by German feedback could therefore explain the tendency for implausible errors to be corrected more often on the delayed test by participants who received German feedback as compared with English feedback. Context could be beneficial in this case because it may provide another chance to infer the word's meaning (Nation, 2001), encourage elaborative processing of the word's meaning (e.g., Cheng & Good, 2009; Craik & Lockhart, 1972; Hulstijn & Laufer, 2001; Laufer & Hulstijn, 2001; Peters et al., 2009), or provide a greater variety of retrieval cues that can facilitate future recall of the word's meaning (e.g., Young & Bellezza, 1982).

The question of how errors affect learning is not yet well understood. These results provide new data on this important question as it pertains to inferring the meanings of unknown foreign words from context, and suggest that two important factors influencing how likely students are to

learn from errors are the nature of the error itself, and the type of feedback that is provided to correct it.

Author Note We are grateful for the help of Mark Looney (Department of World Languages and Cultures, Iowa State University), who assisted with participant recruitment and offered many helpful comments about this research. We thank Hana Ibrik for her assistance in preparing the materials and scoring the data. We also thank Aaron Hurd, Anne Null, Annie Hartman, John Sachs, Jonathan Jackson, and Nichole Kentner for their help with pilot testing.

References

- Carpenter, S. K. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *37*, 1547–1552.
- Cheng, Y.-H., & Good, R. L. (2009). L1 glosses: Effects on EFL learners' reading comprehension and vocabulary retention. *Reading in a Foreign Language*, *21*, 119–142.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*, 671–684.
- Forlano, G., & Hoffman, M. N. H. (1937). Guessing and telling methods in learning words in a foreign language. *Journal of Educational Psychology*, *28*, 632–636.
- Gipe, J. P. (1980). Use of a relevant context helps kids learn new word meanings. *Reading Teacher*, *33*, 398–402.
- Grace, C. A. (1998). Retention of word meanings inferred from context and sentence-level translations: Implications for the design of beginning-level CALL software. *Modern Language Journal*, *82*, 533–544.
- Guthrie, E. R. (1942). Conditioning: A theory of learning in terms of stimulus, response, and association. *Yearbook of the National Society for the Study of Education*, *41*, 17–60.
- Horst, M., Cobb, T., & Meara, P. (1998). Beyond a clockwork orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*, 207–223.
- Huelsen, B., & Metcalfe, J. (in press). Making errors facilitates learning, but learners do not know it. *Memory & Cognition*.
- Hulstijn, J. H. (1993). When do foreign-language learners look up the meaning of unfamiliar words? The influence of task and learner variables. *Modern Language Journal*, *77*, 139–147.
- Hulstijn, J. H., Hollander, M., & Greidanus, T. (1996). Incidental vocabulary learning by advanced foreign language students: The influence of marginal glosses, dictionary use, and reoccurrence of unknown words. *Modern Language Journal*, *80*, 327–339.
- Hulstijn, J. H., & Laufer, B. (2001). Some empirical evidence for the involvement load hypothesis in vocabulary acquisition. *Language Learning*, *51*, 539–558.
- Kang, S. H. K., Pashler, H., Cepeda, N. J., Rohrer, D., Carpenter, S. K., & Mozer, M. C. (2011). Does incorrect guessing impair fact learning? *Journal of Educational Psychology*, *103*, 48–59.
- Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for students of different verbal abilities. *Modern Language Journal*, *78*, 285–299.
- Laufer, B. (2003). Vocabulary acquisition in a second language: Do learners really acquire most vocabulary by reading? Some empirical evidence. *Canadian Modern Language Review*, *59*, 567–587.
- Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*, 1–26.
- Laufer, B., & Sim, D. D. (1985). Taking the easy way out: Non-use and misuse of clues in EFL reading. *English Teaching Forum*, *23*, 7–10.
- Lindgren, A. (1980). *Ich will auch in die Schule gehen*. Hamburg, Germany: Oetinger Verlag.
- Lupescu, S., & Day, R. R. (1993). Reading, dictionaries, and vocabulary learning. *Language Learning*, *43*, 263–287.
- Nation, I. S. P. (2001). *Learning vocabulary in another language*. New York, NY: Cambridge University Press.
- Parlow, J., & Berlyne, D. E. (1971). The effect of prior guessing on incidental learning of verbal associations. *Journal of Structural Learning*, *2*, 55–65.
- Peters, E., Hulstijn, J. H., Sercu, L., & Lutjeharms, M. (2009). Learning L2 German vocabulary through reading: The effect of three enhancement techniques compared. *Language Learning*, *59*, 113–151.
- Pitts, M., White, H., & Krashen, S. (1989). Acquiring second language vocabulary through reading: A replication of the clockwork orange study using second language acquirers. *Reading in a Foreign Language*, *5*, 271–275.
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *Science*, *330*, 335.
- Saragi, T., Nation, I. S. P., & Meister, G. F. (1978). Vocabulary learning and reading. *System*, *6*, 70–78.
- Skinner, B. F. (1958). Teaching machines. *Science*, *128*, 969–977.
- Young, D. R., & Bellezza, F. S. (1982). Encoding variability, memory organization, and the repetition effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *8*, 545–559.