Incidental acquisition of new words during reading in L2: Inference of meaning and its integration in the L2 mental lexicon

Online Supplementary Materials

Overview of the tasks in the whole experimental session

In Task 1, participants read brief passages in which unknown novel words appeared (pseudowords standing for low frequency German words). After each passage, the participants performed a part of an integrated self-paced reading task, whose aim it was to find out whether they had inferred the meaning of the novel word. The passages themselves were presented in two versions, one syntactically simple, and the other syntactically complex. The purpose of this manipulation was to explore whether and how the syntactic complexity of texts affects the inference of meaning of unknown words.

Task 2 was a lexical decision task, which investigated whether participants stored the orthographic form of a new lexical item, whether it could be accessed independently from the context in which the novel word had appeared and whether the success of the word form retention was modulated by the syntactic complexity of the context. This task also provided information about the proportion of recognized forms of the novel words.

In the following semantic priming task (Task 3), the novel words were paired with semantically related existing words. The novel words appeared as primes and the semantically related words as targets. In the unrelated condition, the same targets were paired with semantically unrelated primes. The purpose of this task was to establish whether the newly learned words became integrated into the existing L2 semantic network.

After the three computerized tasks, the participants received sheets with the novel words that were introduced in the first task together with several fillers and were asked to rate them on the VKS (Paribakht & Wesche, 1993, 1996). The results of these self-evaluation tests were compared with the results from the computerized tasks.

Afterwards the participants completed a language history questionnaire that assessed in detail their knowledge of German (Task 5).

In the second to last task the participants were asked to write down the meanings of the low frequency words that had been replaced by pseudowords in the previous tasks.

At the end of the session, the reading span of each participant was measured with a standard computerized version of a reading span test (van den Noort et al., 2008) since earlier evidence has shown that different reading strategies are employed in low-span readers versus high-span readers (King & Just, 1991; Just & Carpenter, 1992; MacDonald, Just, & Carpenter, 1992). However, because no relevant effects of reading span were observed, the test is not reported in this study.

The whole experimental session took approximately 100 minutes, including a few short breaks.

Task 2: Lexical Decision

The purpose of the lexical decision task was to find out whether the participants store the orthographical configuration of the novel words, whether they are able to recognize them without the immediate presence of the context in which they had acquired them and whether these aspects are modulated by the syntactic complexity of the texts in which the novel words appeared.

Method

Materials

The materials for this task consisted of the 20 novel words (pseudowords that had appeared in the texts of the first task), 20 new pseudowords and 20 existing German words that did not appear in the previous or later tasks of the session. Similarly to the novel words, the pseudowords (all phonologically regular) and the existing words also consisted of one or two syllables each (half of them one syllable, the other half two).

Procedure

At the beginning of the session, the participants read the instructions on how to perform the task. They were asked to score as words also those items that they had first seen only in the texts which they had read in the previous task. The task started with six practice items. Afterwards, 60 trials followed in one block that started with three

additional practice items. Each trial started with a fixation point that was presented on the screen for 500 ms. Subsequently the target word/pseudoword appeared on the screen with a maximum duration of 3500 ms and the participants had to make a decision about its lexical status (German word, or not) by pressing the YES or NO button on the keyboard. We expected that participants would recognize about half of the novel words and take the other half for pseudowords, so that the proportion of the yes- and no-answers in the experiment would be approximately half-half. After the participants pressed the response buttons or after the maximum time was up, the stimulus disappeared from the screen and the next trial started after an interstimulus interval (blank screen) of 500 ms. Participants' reaction times were measured and the answers scored for accuracy. The task took approximately three minutes.

The order in which the items appeared on the screen was individually pseudorandomized for each participant. A maximum of two items of the same status (novel word, pseudoword, word) and a maximum of three items of the same syllable number could appear after each other.

Results and Discussion

Single measurements were excluded if they differed more than two standard deviations from each participant's mean.

Table 1. *Mean Reaction Time (in ms), Number of Judgments and Corresponding Percentages for Each Judgement With Respect to Word Types in Experiment 2.*

	Novel words			**Pseudowords**			**Real words**		
	Mean	*N*	*(%)*	*Mean*	*N*	*(%)*	*Mean*	*N*	*(%)*
Incorrect response	1121.0	509	37.4%	1149.9	294	21.6%	1092.9	59	4.3%
Correct response	1043.1	851	62.6%	1088.7	1066	78.4%	751.3	1301	95.7%

Note: „yes" responses (i.e. "it is a word") were scored as correct for the novel words

The ANOVA results showed that the critical groups (correct answers to words and pseudowords and correct and incorrect answers to the novel words) differed significantly ($F1 (3, 66) = 128.34$, $p < .001$; $F2(1,19) = 4.71$, $p < .05$). The post hoc Scheffe tests revealed that participants were fastest in responding "yes" to real words

(751.3 ms). The second fastest responses were the yes-responses to the novel words (1043.1 ms) which differed in both F1 and F2 also from the slowest no-responses to the novel words (1121.0 ms). The no-responses to the pseudowords (1088.7) did not statistically differ from the no-responses to the novel words in F1, and in F2 they did not differ either from the yes- or the no-responses to the novel words.

Participants were thus fastest pressing the yes-button both for the existing words and the novel words and they were slower to refuse both the pseudowords and the novel words which they did not recognize. The pattern of results for the yes- and no-responses to the novel words is thus the same as for the correct responses to the real words and pseudowords, only the reaction times are in general slower for the novel words, probably because their memory traces are still not so firmly established and it takes longer time to access their representations.

The analyses of just the novel words with respect to the factors Complexity, Plausibility and Response Accuracy revealed the pattern summarized in Table 2.

Table 2. *Mean Reaction Times (in ms) and Numbers for Yes- and No-Responses (n) With Their Respective Percentages Regarding the Novel Words in Experiment 2 (N=340).*

		Implausible			**Plausible**		
		Mean	*n/N*	*%*	*Mean*	*n/N*	*%*
No-response	Complex	1099.8	131	38.5%	1204.2	124	36.5%
	Simple	1079.5	124	36.5%	1102.1	130	38.2%
Yes-response	Complex	1017.4	209	61.5%	1079.0	216	63.5%
	Simple	1029.9	216	63.5%	1046.1	210	61.8%

The factors Response Accuracy and Plausibility were significant, other effects and their interactions not. Participants were faster when deciding that a novel word is a word than rejecting it (means 1043 ms vs. 1121 ms; $F1(1, 16) = 5.61$, $p < .05$; $F2(1, 19) = 6.08$, $p < .05$). They were also 46 ms faster when the novel words appeared in the implausible (1047 ms) than in the plausible (1093 ms) condition ($F1(1, 16) = 8.28$, $p < .01$; $F2(1, 19) = 6.27$, $p < .05$).

The latter effect can be explained as reflecting difficulties with integrating the novel word with the semantically incompatible adjective (either on the textual, or on the

lexical level). Participants may have directed more attention to the orthographical form of the novel word in the implausible condition consciously reconsidering whether they e.g. read the novel word correctly after noticing the semantic anomaly. The storage of the orthographical forms of the novel words in this condition might have thus profited from this process. The absence of a complexity effect indicates that syntactic complexity did not affect the storage of the orthographic form of the novel words.

The interesting aspect of the results of the lexical decision task is how unspectacular or ordinary they appear. However, when compared with results of equivalent L1 experiments reported by Bordag, Kirschenbaum, Opitz and Tschirner (in press), the important implications of this "unspectacularity/ordinariness" become more obvious. Contrary to data presented here, the L1 participants did not display the same RT patterns for words and pseudowords on the one hand, and for recognized and unrecognized novel words on the other. While, like the L2 participants in the present experiment, they were faster when responding "yes" to the existing words than when responding "no" to the pseudowords, they were about 100ms slower when accepting the novel words as existing German words than when rejecting them. Moreover, though they recognized about 70% of the novel words in the VKS, they acknowledged the status of existing German words in no more than 40% of the cases in the lexical decision task. The implications of the different results for the L1 and L2 participants in this task, the VKS and the semantic priming task will be discussed later.

Task 3: Vocabulary Knowledge Scale

After the computerized tasks, participants were asked to perform several paper and pencil tasks. We will report the results of the Vocabulary Knowledge Scale (Paribakht & Wesche, 1993, 1996) in this section, because it better fits the logic of the argument. The purpose of this test was to compare individual conscious and explicit knowledge about the novel words with the data from the less explicit computerized tasks and to collect data on how well the participants could recall the meaning of the novel words.

Method

Materials and Procedure

Twenty novel words along with 16 filler items (10 words, six pseudowords, none of them presented in other tasks in the session) were presented in a randomized list.

Participants were asked to judge their individual knowledge of each of the words / pseudowords given a choice of seven possible statements (see Table 3). The completion of this task took approximately 15 minutes.

Results

Altogether 2448 data points were collected from the 68 participants (17 superparticipants). There were 35 missing data points (1.4%) because some participants forgot to fill in one page of the questionnaire.

Participants did not recognize 8.2% of the existing filler words as words and wrongly decided that 11.8% of the filler pseudowords were existing words. Analyses of the novel words (N = 1360, 1285 valid) revealed the following pattern:

Participants rated 39.1% of the novel words as unknown, i.e. about 61% were recognized as words. The comparison with the data from the lexical decision task revealed that participants rated the same novel words as known or unknown in both tasks. As mentioned above, this sharply differs from the results of L1 participants reported by Bordag et al. (in press), who recognized about 70% of the novel words in the VKS, but acknowledged the status of an existing word only to 40% of them in the lexical decision task. Most of the recognized novel words in the present experiment were rated with 2 (“I’ve seen this word, but I don’t know its meaning”; 46.1 %), which means that participants believed that they knew the meaning of only 14.6% of the novel words.

Table 3. *Proportion of Scores for the Novel Words on the VKS.*

		Proportion of Ratings (in %)
1	I’ve never seen this word before.	39.1
2	I’ve seen this word, but I don’t know its meaning.	46.1
3	I think this word means…	9.7
4	I know that this word means…	2.4
5	I know this word very well and can use it in a sentence.	1.2
6	I also know the gender of the word and/or its alternating forms.	1.3
7	I first encountered this word in this experimental session.	45.9

585 (45.5%) of the novel words were (additionally) given a rating of 7 ("I first encountered this word in this experimental session"). This means that participants gave less novel words a rating of 7 than the number they actually recognized (rated 2 or higher). This indicates that they were not always aware that they saw the novel words for the first time in the experimental session. Moreover, this result differs from the L1 data (Bordag et al., in press). The L1 participants were always acutely aware of the fact that they saw the novel words for the first time during the experimental session and gave a rating of 7 to all novel words which they rated higher than 1 (see discussion on the semantic priming task).

Participants incorrectly classified 49 (12.0 %) pseudowords as words which they had encountered in the previous task (rated with a 7).

The ANOVAs with the factors Complexity and Plausibility revealed no significant effects on any of these factors or their interaction (all $F < 1$).

Overall, the results of the VKS show that the method agreed with some of the results from Task 2 with respect to the orthographical form recognition, but could not reliably reveal more specific aspects of the acquisition (plausibility and complexity effects in Task 1). The comparison with the results from Task 1 and from Task 4 also shows that the VKS measure is not sensitive enough to reveal weakly represented newly acquired meanings of the novel words.

Low-frequency word knowledge test

In the last paper and pencil task it was investigated whether participants were familiar with the low frequency words which were replaced by the 20 novel words that were presented within directive contexts (Beck, McKeown, & McCaslin, 1983) in the first part of the task.

All participants were given a list of those 20 low frequency words and were asked to provide a short definition, a synonym and/or a translation for each given word. The results of this test confirmed that the meanings of these low frequency words were mostly unknown to the participants. On average they recognized up to three words, usually words that were cognates with their L1 translations. Participants were thus not simply identifying an already known word and associating it with a new label (learning a synonym), but acquiring a completely new item in their L2 as they typically do during L2 reading. This finding confirms the advantage of using the pseudowords as this way it

could be guaranteed that participants had never encountered the given word form before.

Unmasked priming in the present study

Since the presentation of the primes was unmasked, we cannot completely exclude the possibility that strategic processes were involved in the experiment due to long prime presentation. It must be noted though that although the presentation might seem long (450ms), L2 word recognition generally takes longer, as well the recognition of low-frequency words. When taking these two factors into account, the time available to participants for strategic behaviour substantially shrinks.

In addition to this, it is difficult to imagine how exactly the strategic processes should work that would result in semantic facilitation when primes were familiar words and in semantic inhibition when the primes were novel words. An important aspect in this context is the fact that only responses to targets whose primes were novel words that participants recognized as existing words in the lexical decision task were included in the analyses. Consequently, related primes that were existing words and induced facilitation did not differ in their "word" status from the related novel primes that induced inhibition.

The possibility of strategic behaviour should be further reduced by the fact that (as obvious from the VKS) participants were often not aware that they encountered a particular novel word only in the experimental session – thus, these novel words did not differ for them from other known words that were followed both by words and pseudowords.

It is further not clear how participants could develop strategies involving the meaning of novel words which they did not recall. We agree that the evidence for engagement of novel words in the existing semantic network might be even stronger if masked priming was used. However, we are not sure whether this paradigm could be successfully applied under the given conditions (weak orthographic and semantic representations and L2 learners). We furthermore believe that the circumstances in the present experiment, as mentioned above, favour automatic semantic activation interpretation of the observed effects rather than involvement of strategic influences. In addition, the results of Bordag, Kirschenbaum, Roghan and Tschirner (in progress) clearly showed different priming patterns within the group of novel words, depending

on whether participants could consciously recall their meaning (facilitation) or not (inhibition). The same tendencies were also observed in the present study. The characteristics of the novel words' semantic representations and mechanisms related to their emergence (see below) rather than task dependent strategies thus seem to be a more likely origin of the observed effects.

Summary of relevant results of Bordag et al. (in progress)

Bordag et al. (in progress) explored the role of word form properties (marked vs. unmarked forms) in INTENTIONAL vocabulary acquisition. Rather than deriving meanings from short texts like in the present study, participants studied the meanings of the novel words from definition during an experimental session that resembled the one reported in the present paper. In the VKS, participants recognized about 70% of the novel words and could recall more than 30% of their meanings. It was thus possible to include a new factor in the analyses, which was called Depth of Lexical Knowledge and included two levels based on the results of the VKS: Form (novel words rated by individual participants with 2, i.e. participants recognized the word forms, but could not explicitly recall the meaning) and Meaning (novel words rated by individual participants with 3 - 6, i.e. participants recognized the word form and could recall the correct meaning). The 2 x 2 ANOVA with factors Relatedness and Depth of Lexical Knowledge revealed a significant interaction between the two factors. Whereas the interaction between the semantic representations of novel words whose meanings participants could explicitly recall and older representations lead to facilitation (as observed for known words in the present study), the interaction between the meanings of novel words that participants could not explicitly recall and earlier established related semantic representations lead to inhibition as in the present study. Strength of semantic representation and/or the ability to explicitly recall the meaning thus seems to be a decisive factor for either facilitatory or inhibitory semantic priming effects to appear.

Since incidental acquisition leads to smaller vocabulary gains than intentional acquisition, as shown in Bordag et al. (in progress), and since the proportion of novel words whose meaning participants could recall in the present study is very low (14.8%), analyses of the present data with the factor Depth of Lexical Knowledge, as in Bordag et al. (in progress), are deemed to be insignificant due to a lack of statistical power. However, the numerical tendencies show the same pattern of results as reported by

Bordag et al. (in progress) and the interaction between Relatedness and Depth of Lexical Knowledge actually even reaches significance in F2: F2(1, 65) = 7.6, p = .007. Whereas novel words whose meanings participants can recall induce facilitation, those whose meaning participants cannot recall lead to inhibition (see the following Table 4):

Table 4. *Mean Reaction Times (in ms), Number of Analyzed Data Points, and Their Respective Percentage per Experimental Condition, Taking into Account the Factor Depth of Lexical Knowledge.*

Depth of Lexical Knowledge	**Priming Condition**	**Semantic Relation**						**Diff**
		Related			**Unrelated**			
		Mean	*N*	*%*	*Mean*	*N*	*%*	
Form	Semantic	719.7	543	41.1	701.0	553	41.9	**-14.5**
	Sem. & context.	701.0	554	42.0	680.3	553	41.9	**-18.2**
Meaning	Semantic	652.5	182	13.8	675.9	187	14.2	**+24.0**
	Sem. & context.	650.0	188	14.2	658.4	180	13.6	**+11.5**

References

Beck, I. L., McKeown, M. G., & McCaslin, E.S. (1983). Vocabulary Development: All Contexts Are Not Created Equal. *Elemetary School Journal 83*(3), 177–181.

Bordag, D., Kirschenbaum, A., Opitz, A., & Tschirner, E. (unpublished manuscript): To store, or not and if so, where? An experimental study on incidental vocabulary acquisition by adult native speakers of German. In V. Torrens & L. Escobar (eds.), *The Processing of Lexicon and Morphosyntax.* Newcastle: Cambridge Scholars Publishing.

Bordag, D., Kirschenbaum, A., Roghan, M., & Tschirner, E. (unpublished manuscript): The influence of form (a)typicality on incidental and intentional vocabulary acquisition.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149. doi: 10.1037/0033-295X.99.1.122

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: the role of working memory. *Journal of Memory and Language, 30,* 580–602.

MacDonald, M., Just, M., & Carpenter, P. (1992). Working memory constraints on the

processing of syntactic ambiguity. *Cognitive Psychology, 24*, 56–98.

Paribakht, T.S., & Wesche, E. (1993). The relationship between reading comprehension and second language development in a comprehensionbased ESL programm. *TESL Canada Journal, 11*(1), 9-29.

Paribakht, T. S., & Wesche, M. (1996). Enhancing vocabulary acquisition through reading: A hierarchy of text-related exercise-types. *The Canadian Modern Language Review, 52*(2), 155-178.

van den Noort, M., Specht, K., Rimol, L. M., Ersland, L., & Hugdahl, K. (2008). A new verbal reports fMRI dichotic listening paradigm for studies of hemispheric asymmetry. *Neuroimage, 40*, 902–911. doi: 10.1016/j.neuroimage.