Introduction

A Practical Outline of Intentional Learning

It is generally possible, over a given period of time, for post-early-childhood learners to acquire more L2 vocabulary intentionally than incidentally (Barcroft, 2015; Hulstijn, 2003; Laufer, 2005). Thus, one might expect there to be numerous pertinent, in-depth discussions about what intentional learning is and about how it differs from incidental learning. However, such discussions are scarce and their portrayals of intentional learning are indistinct. Indeed, it seems that in cognitive science there is no current theory of learning that includes the presence or absence of an intention to learn as a construct or a critical element (cf., Hulstijn, 2003, p. 356). That said, few if any modern authorities doubt that people can make decisions about what to try to learn, how to try, when to start, and when to quit (e.g., Bjork, Dunlosky, & Kornell, 2013). This must be one reason why, theory or no theory, the notion of intentional learning is one that L2 researchers can hardly do without.

Going by how intentional learning is referred to in the literature on L2 vocabulary learning, it falls under the heading of language-focused learning, one of Nation’s (2013) “four strands” of a well-balanced program for learning an additional language. Authorities on memory such as Roediger (2008) see intentional learning as separable into intentional encoding and intentional retrieval. In experimental psychology as a whole, it is operationalized as learning that happens when those who learn are aware that there will be a test of retention, whereas incidental learning is deemed to occur when those who learn are not aware that a retention test is coming (Baddeley, Eysenck, & Anderson, 2009). However, this sharp distinction poorly reflects the full range of situations in which intentional learning in some sense evidently occurs (Barcroft, 2015; Hulstijn, 2001, 2003; McLaughlin, 1965; Postman, 1964). One fact left out of account by the operational definition of intentional learning is that in a school setting or in the context of an experiment, learners may well suppose that a test will follow an episode of vocabulary study even when no forewarning of a retrieval test is given. (Thus, researchers aiming to maintain conditions of incidental learning sometimes try to forestall such suppositions by feeding participants a cover story.) More
Intentional L2 Vocabulary Learning

fundamentally, the operational definition takes no account of the likelihood that individuals can have intentions to learn that are independent of any belief on their part that a researcher, teacher, or other authority is going to test them (e.g., Barcroft, 2015). Another difficulty for anyone seeking to sharply delineate intentional learning is the evidence that a practice that is ordinarily associated with intentional learning – namely, using word cards to memorize L2 forms and form-meaning mappings – can result in a level of fluent accessibility in memory which is typical of what results from long-term incidental learning (Elgort, 2011). I should mention that there is evidence from a study involving electro-encephalogram recordings that incidental and intentional learning of L1 words result in different modifications of functional brain networks associated with verbal episodic memory (Kuhnert et al., 2013), but it is not clear what this might imply for learning L2 words or for being able to retrieve their meanings and forms from a memory store which supports everyday language use. Moreover, weighing against speculation that incidental and intentional learning may be qualitatively different activities, there is a wide range of evidence that, in terms of behavior, the difference between intentional and incidental learning is a matter of degree (Postman, 1964; McLaughlin, 1965). Importantly, there are circumstances in which learning appears to be wholly unaffected by the level of an individual’s intention to learn; and the same can be said of levels of desire and of motivation. Rather than being crucial factors in learning these are factors that may influence learning indirectly, if at all (Baddeley et al., 2009). What is crucial is how information is processed (Baddeley et al., 2009). For the time being, then, it seems best to think of intentional learning not as a type or even a mode of learning per se but to think of intention as a possible condition on learning (cf., Roediger, 2008) that may be absent or present to a greater or lesser degree. Finally, it has been remarked (Laufer, 2005) that the idea of intentional learning of vocabulary assumes that one can learn vocabulary essentially as one can learn nonlinguistic skills and bodies of knowledge in the sense that (1) lexical items that have been learned can be part of one’s declarative knowledge, (2) knowledge of these items can also become proceduralized, whereby they can be used and understood without conscious attention and control, and (3), ultimately their forms may become automatically and instantaneously available when needed.

Viewing intentional learning now from the standpoint of someone who reads articles reporting quantitative investigations of L2 vocabulary learning, it is noteworthy that one can decide whether or not they have to do with intentional learning with widely varying degrees of confidence. While it seems reasonable to assume that the clearest cases of intentional learning do in fact feature forewarning of a retrieval test, additional indicators are needed.1 An especially useful one is whether participants in a study were asked to learn a specified set of items (Barcroft, 2015; McLaughlin, 1965; Postman, 1964).2 Two other indicators are whether learners received explicit instructions about how to engage with, or process, the to-be-learned items (e.g., McLaughlin, 1965; Postman, 1964) and whether learners followed the instructions with signs of awareness that the main goal was to learn vocabulary rather than, say, just to read and understand a story. In other words, a situation of intentional learning tends to differ from a situation of incidental learning in that the former involves a high degree of focus on vocabulary learning, rather than “on other possible tasks or other possible stimuli in the learner’s environment” (Barcroft, 2015, p. 51). To be more specific here, an instruction to count the number of syllables in specified items is likely to lead to intentional processing of the items’ forms, while the instruction to use each targeted item in a meaningful written sentence is likely to result primarily in the intentional processing of meanings (Barcroft, 2015). To continue, the presentation of targeted items in a situation of intentional learning tends to follow a pattern such that the items appear a predetermined number of
times and for predetermined durations (Barcroft, 2015). Finally, to come at the matter from two other directions, intentional and incidental learning have been found to show differences of degree with respect to certain variables. For example, whereas the objective, real world frequency of (partly) unfamiliar lexical items is a key variable in incidental vocabulary learning, it is apparently of minor or even negligible importance in intentional vocabulary learning (de Groot & Keijzer, 2000; Lindstromberg & Eyckmans, 2019). Also, the superiority of intentional over incidental vocabulary learning that is seen in tests of retrieval may disappear or even be reversed in a test of recognition (Eagle & Leiter, 1964). It seems relevant to add that the replicability of many common findings in the experimental study of intentional learning may depend on what type of test or experimental design is used (Roediger, 2008). For example, it is commonly found that “generation” (or production) of to-be-learned information promotes learning better than just “receiving” (e.g., reading) it; however, a design change can cause the generation effect to disappear or even reverse (Roediger, 2008).

The Scope and Aims of Discussion

Because various practicalities of intentional L2 vocabulary learning are dealt with in other chapters of this book, the main aims of this chapter are few in number. Indeed, the main aim of describing intentional learning has already been partly accomplished. The additional principal aims are to elaborate on certain crucial matters of background and then to call attention to a few of the complexities in carrying out and making sense of relevant empirical studies. These studies are ones that involve what Laufer (2005, p. 223) has called “pure focus on forms” (my emphasis), and they have to do with vocabulary learning in planned episodes occurring neither within nor in close relation to a communicative task of the sort advocated by recent proponents of task-based learning (e.g., Ellis, Basturkmen, & Loewen, 2002).³ This category of studies overlaps considerably with Barcroft’s category of studies concerning “intentional learning and direct instruction”, which he places at the most intentional end of the “incidental-intentional continuum” (2015, p. 48). It is this end of the continuum that we focus on in this chapter. (For discussion of pedagogical techniques which evidently fall nearer to the middle of the continuum, e.g., provision of input enhancement and textual glosses (Barcroft, 2015), see Pellicer-Sánchez, this volume.)

Critical Issues and Topics

Although matters important for understanding a discussion of intentional learning L2 vocabulary have been touched on in other chapters, it may be helpful to take a longer look at several of them.⁴

Retrieval Practice and the Testing Effect

In an authoritative appraisal of ten much-studied and often-recommended practices and techniques for the intentional learning of information of various kinds, Dunlosky, Rawson, Marsh, Nathan, & Willingham (2013) concluded that only two are of high utility. One of these, distributed practice, is discussed in Chapter 20 (see Nakata, this volume). The other is retrieval practice (aka practice testing), which Dunlosky et al. (2013, p. 6) define as “self-testing or taking practice tests over to-be-learned material”. The rationale for retrieval practice is based on the testing effect, which is the strong and consistent finding that information that is actively recalled becomes more retrievable subsequently than would be the
Intentional L2 Vocabulary Learning

case without a test or if the test were replaced by an equally long session of restudying (e.g., Karpicke & Roediger, 2008; Roediger & Butler, 2011; Karpicke, 2017). An important qualification here is that although a retrieval test generally turns out to be more beneficial in the long-term than an equally long-lasting session of restudying, restudying tends to appear most beneficial in the short term. A partial explanation for this common finding is that a test of retrieval strengthens relevant memory traces more durably than restudying does because it is more focused.

To say more about retrieval, a recent hypothesis (Kornell, Klein, & Rawson, 2015) sees it as a two-stage process. First, an individual becomes aware of a “retrieval cue”, such as a question or a gap in a sentence, and the cue triggers a mental search for an appropriate response. Second, the individual either retrieves the response or does not retrieve it but luckily is given the response “from the outside” (perhaps in the form of feedback). Either way, appropriate connections in memory are strengthened; and, in some cases, inappropriate connections are weakened. The key conclusion is that what matters is not retrieval in and of itself but rather the processing that goes on during the search for the response and the processing that goes on after the response becomes available. To elaborate, it is during the first episode of processing that the memory traces of related information are activated and it is during the second episode that connections among the memory traces for the cue, for the related information, and for the response become stronger. Because difficult retrieval entails more processing, tests which make retrieval easy tend to be less effective in the long term than tests that make retrieval effortful. For example, tests that require retrieval from working memory, as opposed to secondary (or long term) memory, are likely to bring negligible benefits (Roediger, 2008). To ensure that retrieval draws on secondary memory and is therefore sufficiently effortful, a learner can seek (or be given) feedback only after a delay. Indeed, it has been proposed more generally that there is such a thing as “desirable difficulty” in the learning of anything (Bjork & Kroll, 2015). Kornell et al. (2015, p. 291) have gone so far as to say that because learners profit least from retrieving information about which they are relatively sure, “retrieval success may be a sign that relatively little learning is occurring and that one should have waited longer before attempting to retrieve” (Kornell et al., 2015, p. 291). As to feedback, it should be stressed that it is particularly necessary that feedback be given after tests which expose learners to incorrect information, as recognition and multiple choice tests do (Roediger & Butler, 2011). If feedback is not given, then the expectation should be that rates of successful retrieval will be reasonably high.

Two types of tests known to engender the testing effect are tests of free recall and tests of cued recall (where a cue for retrieving an L2 word might be an L1 equivalent or the L2 word’s first letter). Also effective are variants of cued recall in which learners produce an L2 word or phrase in response to a question or a gap in a context. Crucially, the testing effect can eventuate not just from formal classroom tests but from retrieval exercises in general, including retrieval activities with a game-like character. Thus, incorporating more testing practice into a syllabus or study plan does not mean that instruction or autonomous study must be conspicuously laborious and serious. A possible (and possibly culture-specific) obstacle to implementing a regime of markedly test-like learning events is suggested by reports that American learners, for instance, are likely to doubt that being tested is a good way to learn (Dunlosky et al., 2013; Karpicke & Roediger, 2008). Also, it is easy to imagine that young learners may wilt under a regime of retrieval practice unless a proportion of the tests are somehow made fun to do. Finally, alleged differences in personal learning styles do not constitute an impediment to widespread implementation
of good learning practices such as retrieval practice (Arbuthnott & Krätzig, 2015; Pashler, McDaniel, Rohrer, & Bjork, 2008).

Levels of Processing (LOP) Theory and the Elaborative Processing of Lexical Items

The notion of deep, rich, or elaborative mental processing of, say, L2 vocabulary stems substantially from Levels of Processing Theory (LOP) (e.g., Craik, 2002; Craik & Lockhart, 1972; Craik & Tulving, 1975). Key ideas in LOP theory as set out by Craik and Tulving (1975) are that the durability of a memory trace is a positive function of degree of “semantic involvement” (i.e., engagement with meaning) (p. 268) and that memory performance depends on the elaborateness of the final encoding of lexical information.  

Semantic Elaboration in General and Image-Based Elaboration in Particular

With reference to a given lexical item, semantic elaboration is an umbrella term in LOP theory for mental operations resulting in a more detailed mental representation of the meaning of a given lexical meaning item. This may include the addition of neural-connections to memory traces for meanings of other lexical items. An important general type of semantic elaboration is the integration of the meaning of a newly or partly learned vocabulary item into the mental lexicon (e.g., Thompson, 1987). A specific case of semantic elaboration is mapping the meaning and form of an L2 target to the meaning and form of an L1 cognate.

An additional type of semantic elaboration is entailed by Dual Coding Theory (Paivio, 1986; Paivio & Desrochers, 1979; Paivio, Walsh, & Bons, 1994; see also Holcomb, Anderson, Kounios, & West, 1999; and, with respect to multiword expressions, Li, 2009). This theory holds that many items of vocabulary are encoded (or represented) in memory in one or both of two ways. One type of encoding is iconic in character and one is purely verbal in character. The former type of encoding supports lexical meanings that include images – visual and sensori-motoric ones, for example. Because imageability is highly correlated with concreteness (generally, $r \approx .84$), abstract vocabulary – unless it is quite emotive – may be encoded in memory with few if any associated images (Dellantonio, Mulatti, Pastore, & Job, 2014). It is well documented that imageable L2 words are comparatively easy to acquire in situations of intentional learning. Steinel, Hulstijn, and Steinel (2007) found that the same is true for L2 idioms. Dual Coding Theory suggests that encoding imageable words in both verbal and iconic channels rather than in just one channel, enhances learnability.

A number of researchers have attempted to confirm that learners’ ability to learn and durably remember L2 expressions is enhanced by tasks intended to promote imagistic encoding. For example, Boers (2001), Boers, Demecheleer, and Eyckmans (2004), and Boers, Eyckmans, and Stengers (2007) have reported good results for an approach to learning L2 English figurative idioms which involves exploring or speculating about the domain of activity in which targeted idioms are, or were, used literally. A key rationale is that such reflection is likely to feature the generation of mental images. For additional discussion, see Boers and Lindstromberg (2008, 2009).

The (Elaborative) Processing of Lexical Forms

Although LOP theory recognizes that learners can process orthographic and phonological forms, past versions of the theory heavily prioritized meaning (i.e., semantic) processing. This probably accounts in part for the fact that L2 researchers have paid much more attention
Intentional L2 Vocabulary Learning

to (elaborative) meaning processing than to (elaborative) form processing. Let’s consider a recent theory that takes a balanced view of these two types of processing along with a posited third type of processing, the processing of form-meaning mappings.

**Barcroft’s “Type of Processing – Resource Allocation” Model**

Barcroft and colleagues (e.g., Barcroft, 2002, 2006, 2007, 2015; Barcroft, Sommers, and Sunderman, 2011) have presented persuasive evidence that LOP theory requires important qualifications with respect to L2 vocabulary learning (see also Roediger, 2008). To begin at the beginning, Craik and Tulving (1975) described a within-subjects experiment (their experiment number 4) which involved the free recall of L1 concrete nouns studied in a situation of intentional learning. Of the three experimental conditions, one called for processing of orthographic form (e.g., participants were asked whether a targeted word was shown in capital letters), one called for processing of phonological form (i.e., participants were asked whether a targeted word rhymes with a certain other word), and one called for processing of meaning. In the latter condition participants were asked to judge whether or not a targeted word fits thematically in a given gapped sentence. Subsequent free recall of words encountered in the meaning-processing condition was found to be superior to free recall of words encountered in the two form processing conditions. Since Craik and Tulving (1975) equated semantic processing with deep processing and processing of form with shallow (i.e., relatively superficial) processing, they concluded from the results of this experiment, and from similar ones, that semantic processing promotes word learning more effectively than processing of form does. Barcroft (e.g., 2015), however, has argued that the relative effectiveness of semantic processing and form processing (aka, structure processing) crucially depends on how much there is to be learned about a given L2 item. That is, with respect to a given learner, questions such as the following are germane: Does the item have both an unfamiliar form and an unfamiliar meaning, that is to say, a meaning which is not expressed by a lexeme in L1? Or is only the form of the item unfamiliar, because its meaning is known from L1? A key to the relevance of these and certain other questions is Barcroft’s resource depletion hypothesis, which holds that there is a limit to the processing capacity available to a given learner in a given situation with respect to a lexical item encountered in input. So, for instance, a learner who is intensively processing meaning may have little or no capacity left over for the processing of form (and vice versa). This hypothesis is an element of Barcroft’s Type of Processing – Resource Allocation (TOPRA) model, the gist of which is given by the following three statements. First, processing of form, processing of meaning, and processing of form-meaning mappings can each operate largely independently of each other. Second, “increases in one of these three types of processing . . . increases processing and learning for the type of processing in question, which is reflected in increased learning for the component of vocabulary learning in question” (2015, p. 90). However, “increases in one of these three types of processing . . . can decrease available processing resources for the other types of processing, which is reflected in the relative amount of learning associated with the different types of processing and learning counterparts in question” (p. 90).

One implication of the TOPRA model is that a task which focuses learners’ attention on meaning can lead to comparatively good results in a test of recall (such as the test figuring in Craik and Tulving’s experiment 4 referred to earlier) if, but only if, the learners already have good knowledge of the forms of the to-be-recalled items. (It should be noted that to-be-learned words in Craik and Tulving’s experiment were common L1 nouns). If, on the other hand, learners do not already know the forms of the to-be-learned items to a substantial
degree, then engaging in intensive processing of meanings will leave them with insufficient capacity to process the forms. These learners are likely to do worse in a test of recall than learners who have engaged in a task involving supposedly shallower processing of form. As may be evident, and as pointed out by Barcroft (2015), his TOPRA model is consistent with Morris, Bransford, and Franks’s (1977) influential model of transfer-appropriate learning, which holds in essence that recollection will be best when the processes that take place during encoding match the processes that take place during retrieval.

For a better idea of what TOPRA means, let’s consider aspects of a within-subjects study summarized in detail by Barcroft (2015). That researcher began by asking Anglophone learners of Spanish to try to learn 24 unfamiliar Spanish words that they were soon going to encounter. In the encounter phase itself, each word was matched with an illustrative picture. Eight words were encountered in a meaning-processing condition where the task was to give each word a pleasantness rating. Eight other words were encountered in a form processing condition where the task was to count the letters in each word. For the remaining eight words, participants were given no instruction beyond the common instruction that they should try to learn these words. The study phase was followed by two posttests of free recall and one posttest of cued recall. In the first of these tests, participants were asked to write down as many of the 24 to-be-learned Spanish words as they could remember. In the second, they were asked to write down all the English equivalents that they could remember. In the third, they were shown a picture corresponding to each Spanish word and asked to write down the Spanish words in question. On the first posttest, which probed participants’ ability to recall the forms of targeted words, the average score for the words studied in the form processing condition exceeded the average score for the words studied in the meaning-processing condition. The results of the second test (which probed participants’ ability to recall the meanings of targeted words) were the reverse of those for the first test. Apparently, and as predicted by the TOPRA model, participants who were encouraged to process meaning experienced reduced capacity to process form whereas participants who were encouraged to process form experienced reduced capacity to process meaning. Finally, the highest average scores on all three tests were for the eight words that participants were left to study however they wished. Barcroft’s (2015) interpretation of this outcome is that neither of the specific processing tasks took account of the fact that participants were being asked to learn new forms and new form-meaning mappings all at the same time, with the result that both tasks depleted participants’ capacity to encode the targeted words in an optimal manner.

In another TOPRA-related within-subjects investigation, Barcroft (2003) investigated what happens when learners are called on to answer questions about the meanings of previously unfamiliar words that they are in the midst of learning. For 12 of the 24 targeted words, the learners’ task was to consider questions about each word’s meaning (e.g., In what way can this object be used?) whereas their task for the other 12 words was just to try to learn them somehow. Scores in an immediate and in a ten-minute delayed posttest of picture-to-target cued recall were lowest in the question condition, that is, the condition which encouraged focus on the meanings of the new L2 words. Even greater negative effects of asking learners to engage in semantic processing of novel L2 words have been observed in various studies in which the semantic processing task was to use each targeted word in a written sentence illustrating its meaning (for summaries and references, see Barcroft, 2015). The clear implication of such results is that a task that tends to induce a specific type of lexical processing can be productive or detrimental depending on circumstances (cf., Roediger, 2008). For instance, a task which is oriented toward meaning processing can help learners remember L2 forms that they have already acquired, but it can also sharply reduce learners’ ability to remember word
Intentional L2 Vocabulary Learning

forms that are new to them. To state the obvious, this is important because a word whose form cannot be recalled is a word that cannot be spontaneously used in communication.

Finally, to touch on a matter of research methodology, recall that in the TOPRA-related experimental studies just summarized the baseline condition was defined by the instruction to learners that they should “do their best to learn the target words in question” (Barcroft, 2015, p. 77). Barcroft’s rationale for having such a condition is that the effects of a task which promotes a specific type of lexical processing (e.g., semantic processing) should ideally be compared with a task that does not promote a specific type of processing. A possible disadvantage of having a just-do-your-best baseline condition is the difficulty of determining what strategies participating learners applied and what sort(s) of lexical processing they engaged in. It may therefore be argued that it is better to have at least one condition in which learners are set a comparison task that they are all likely to carry out in more or less the same way – a sensible proviso being, of course, that this task should have effects that can be estimated from prior research. So far as I know, the L2 research community has not reached a consensus about this issue.

A Case Study of Complexities Encountered in Research: The Keyword Method

Many streams of research have cast light on the effectiveness of types of tasks (methods, techniques, and so on) that play, or may play, a positive role in intentional L2 vocabulary learning. This section reviews one of these streams, the one comprising investigations of the keyword method.

For purposes of illustration, the keyword method seemed a good choice. First, use of this method is an extremely clear case of intentional L2 vocabulary learning. Second, the keyword method has figured in well over 100 published studies since it was introduced in the 1970s (e.g., Atkinson, 1975; Atkinson & Raugh, 1975; Ellis & Beaton, 1993; Sagarra & Alba, 2006; Shapiro & Waters, 2005) and in a number of unpublished theses and dissertations as well (e.g., Hauptmann, 2004). (For additional references, see Barcroft, 2015; Hauptmann, 2004; Nation, 2013; Wei, 2015; and www.victoria.ac.nz/lals/resources/vocrefs/vocref-20.) Third, consideration of the literature on the keyword method provides a background against which certain complications may be easier to discern.

In what seems to be the commonest version of the keyword (or linkword) method a researcher, teacher, or a learner chooses a familiar, and highly concrete L1 word, the keyword, as a mediator between a targeted L2 word and an L1 translation. For optimum recollection of form as well as meaning, the keyword should be phonologically similar to (part of) the L2 target. The next step is to think of an image which (1) is based on the meaning of the keyword and (2) involves both the L2 target and the L1 equivalent. Shapiro and Waters (2005) gave the example of English steel as an L1 keyword for learning L2 French stylo (“pen”) and the mediatory image of a large pen made of a steel girder. Shapiro and Waters also described how, for hopefully enhanced mnemonic effect, this image can be played out mentally, like a movie clip.

Many results have indicated that use of the keyword method can result in impressive rates of receptive word learning. For example, Atkinson and Raugh (1975) found that the participants in their study learned 120 words in three days. The effectiveness of this method has been attributed to various combinations of the following partly overlapping factors: (1) elaborative imagistic mental processing of word meanings (e.g., Ellis and Beaton, 1993; Baddeley et al., 2009), (2) increased distinctiveness of the encoded information (Marschark & Surian, 1989), (3) integration of new knowledge with old knowledge (Thompson, 1987),
(4) provision of retrieval cues (Thompson, 1987), and (5) encouragement of effortful processing (Shapiro & Waters, 2005).

Notwithstanding the mountain of results already accumulated, the keyword method continues to be investigated by L2 researchers who are, for instance, interested in assessing its effectiveness compared to that of other methods of intentional vocabulary learning. Not all results are wholly positive. In particular, some reports have noted limitations with respect to type of learner. Campos, Amor, and González (2004) concluded from their results that the keyword method is most suited to learners with high imaging capacity. Wei (2015) concluded from her results and from results of five earlier studies that the keyword method may not be worthwhile for “experienced and intellectually mature learners” (p. 60). An additional, and perhaps more serious concern is that when words are learned using the keyword method, access to their stored meanings is likely to be less direct and slower to negotiate than is so for words learned, say, by rote association of meanings and forms (Barcroft, Sommers, & Sunderland, 2011). In any case, effectiveness is not the only virtue that a method should have: Efficiency, practicality, versatility, and appeal to learners and teachers matter too. Taking such matters into consideration, Dunlosky et al. (2013) concluded that the keyword method is of “limited utility” because (1) it is not applicable to items of low imageability (also, the method seems inapplicable to MWEs [Boers & Lindstromberg, 2009]), (2) the time needed for training and for keyword generation can be better spent, and (3) it is not only insufficiently clear that the keyword method results in durable learning but there is even some evidence that it leads to accelerated forgetting. To elaborate about the last point, in many studies of the keyword method all targeted words were tested in two posttests of retrieval, one immediate and one delayed, and these tests were taken by all participants. In such a design, superior delayed posttest results in the keyword condition could be an artifact of experimental design because if more of the targeted words are retrieved in the keyword condition in the immediate posttest than are retrieved in the comparison condition(s), which is a typical outcome, then later retrieval of targeted items in the keyword condition would be likely to benefit disproportionally from retrieval practice. Having identified this as a potential problem, Wang, Thomas, and colleagues got around it by conducting a series of experiments in which all participants experienced the treatment to which they were randomly assigned, but only half of each treatment group took the immediate test (Wang & Thomas, 1995; Wang, Thomas, & Ouellette, 1992; Wang, Thomas, Inzana, & Primicerio, 1993). The delayed test was taken by the remaining participants in each treatment group. Each of these experiments included a condition in which the researchers provided the keywords (but participants were asked to think up their own image). There was also a comparison condition, which was either rote learning of a list of L2–L1 equivalents or study of L2 words in meaningful sentences; and some experiments included a third condition in which participants were asked to choose keywords on their own. For example, Wang et al. (1993) randomly assigned 89 participants to two groups and informed them of eventual testing: $n_{\text{Words}} = 24$ (L2 tagalog); 2 conditions $\times$ 2 test times. The immediate and delayed posttests were of two types: cued recall (cue = L2 word; response = L1 meaning) and matching L2 words with L1 equivalents. On the immediate posttest of cued recall, the mean scores of the keyword and rote learners were virtually identical ($\approx 79\%$ successes). By the time of a delayed posttest only two days later, the keyword learners had forgotten the meanings of nearly double the number of targeted words that the rote learners had. Relative performances on the delayed matching test were similar. Wang and Thomas (1995) replicated these findings using a comparison condition in which learners studied L2 targets presented in meaningful sentences (see also Thomas & Wang, 1996). One of the four experiments reported by Wang et al. (1992) had the standard design.
in which immediate and delayed posttests were taken by the same learners: Only in this case was the initial superiority of the keyword learners preserved in the delayed test. With respect to the steep forgetting curves they observed for the keyword learners in the three of their experiments that controlled for effects of retrieval practice on the immediate posttest, Wang et al. (1992) speculated that pre-experimental associations to keywords interfere with retrieval of the interactive images formed during the treatment to the extent that during the delayed test, even if learners recall their keywords and keyword images, it may be difficult for them to remember which of the multiple elements in the image went with the keyword. Additionally, it is not clear that the keyword method results in any substantial processing of the form of a L2 target or even of its meaning since the method may well induce learners to devote too much attention to the L1 keyword (cf., Barcroft, 2015) and to the linking image. This could be a problem since, after all, the phonological form of a keyword is rarely identical to that of the L2 target, and the linking image is bound to include elements that have nothing to do with the meaning of the target. As might be expected, then, the wider literature on the keyword method is peppered with accounts of how learners and teachers have found it unappealing (e.g., Hauptmann, 2004; Hulstijn, 1997).

**Future Directions**

A major aim of experimental research is theory development. Accordingly, experimental studies of intentional L2 vocabulary learning do not necessarily involve tasks that teachers and learners would regard as useful. This probably accounts for the rather large amount of attention that researchers have paid to the keyword method and to certain other time-consuming and generally problematic procedures, such as composing, and then writing out, meaningful sentences that use targeted expressions. Although theory development is desirable, it would be good if experimental studies nevertheless focused more often on tasks that are feasible and useful outside of an experimental context. The possibility of change in this direction may depend on prior refinement of contemporary theory. It seems to me, for example, that the TOPRA model – which has been a very beneficial spur to research – lacks detail in at least two regards. Specifically, it would be good to know how it is that mapping processing differs from simultaneous processing of form and meaning. As to the disadvantage of trying to carry out two types of processing at the same time, it can be asked how long one must wait after finishing one type of processing before beginning another. If wait-time can be very brief, then the resource depletion hypothesis may not be a formidable factor in task design. The relevance of these issues is exemplified by experimental studies of the form-focused practice of “word writing”, that is, writing to-be-learned L2 words out by hand. While results of some studies indicate that this practice is not beneficial (e.g., Barcroft, 2006, 2007), results of other studies indicate the contrary. Specifically, Thomas & Dieter (1987); Candra, Elgort, Deconinck, & Eyckmans (2017); Elgort, Candra, Eyckmans, Boutorwick, & Brysbaert (2018); and Webb & Piasecki (2018) found statistically significant benefits of word writing in situations where form and meaning were processed. However, it may be necessary for learners to work at their own pace rather than under time pressure (Webb & Piasecki, 2018).

Finally, to consider intentional L2 vocabulary learning in a general way, there is a great deal of evidence that it can be highly fruitful to approach vocabulary learning in an organized manner with a readiness to invest time and effort. Engaging in retrieval practice is recommendable, especially if it includes plenty of tests of ability to recall spoken and written forms as well as meanings. It would be interesting to know whether cultural background is an important variable in how likely learners are to tackle vocabulary learning in the spirit just indicated.
Further Reading


This recent major survey of evidence for the effectiveness of retrieval practice follows on from Roediger and Butler (2011), which I also recommend.


Although written for specialists, this article provides a clear and detailed account of the place of retrieval in contemporary theory of learning and memory.


This is a master class on thinking about learning vocabulary in another language.


Here Roediger explains why there are no known laws of learning and memory.

Related Topics

Incidental learning, retrieval practice, desirable difficulty, experimental design, levels of processing, (elaborative) processing of meaning, (elaborative) processing of form, dual processing, Type of Processing – Resource Allocation (TOPRA) model, keyword method

Notes

1 It appears that a sizeable minority of L2 research reports include no mention of whether learners were forewarned of tests or not.

2 In reading for this chapter I came across no empirical studies in which learners were asked to pre-specify their own items to target.

3 A useful reminder from Laufer (2005) is that the terms *focus on form* and *focus on forms* (which come to us from the literature of task-based learning) disguise the fact that both orientations concern meaning as well as form.

4 Note that the Involvement Load Hypothesis is not discussed in this chapter since it was conceived with incidental learning in mind. In any case, the construct of “search” is scarcely relevant in intentional learning.

5 For Baddeley et al. (2009) LOP is a hypothesis; for Roediger (2008) it is a frequently observed effect for which a theory is lacking.

6 The experimental design of Wang and Thomas seems to have been radically underused by subsequent L2 researchers with respect to interventions whose effects might be strong initially but might dissipate relatively quickly.

7 Baddeley et al. (2009) discuss strengths and limitations of some other word learning mnemonic techniques.

References


Intentional L2 Vocabulary Learning


251


Intentional L2 Vocabulary Learning


