Language Teaching (2022), **55**, 1–23
doi:10.1017/S0261444821000252

CAMBRIDGE UNIVERSITY PRESS

STATE-OF-THE-ART ARTICLE

Glossing and vocabulary learning

Frank Boers

The University of Western Ontario, London, Canada
*Corresponding author. Email: fboers@uwo.ca

Abstract

This article offers a critical review of research on the use of glossing and its contribution to second language (L2) vocabulary acquisition. Discussion topics include the complexity of estimating the effectiveness of glossing relative to reading non-glossed texts, the quest for optimal implementations of glossing, issues of ecological validity, and ambiguity around the nature of vocabulary learning from glosses. The general conclusion is that, despite the substantial number of research studies on this subject, many questions remain to which only tentative and provisional answers are currently available. This is partly owing to the wide diversity in research designs across studies and the lack of transparency of many research reports. Suggestions are made for further research on glossing with a view to enabling future systematic reviews to produce more nuanced answers and more informed recommendations for the design of L2 reading materials.

1. Introduction

Glosses (also known as annotations) are clarifications that are added to texts to support comprehension. They may be placed in the margin of the text close to the line where the item of interest occurs, at the bottom of the page, or as a glossary at the end of the text. Sometimes glosses are inserted within the body of the text, adjacent to the item they are meant to clarify. In the case of computer-mediated reading, the glosses can be provided through hyperlinks, where readers know they can access information by mouse-clicking on the item or by moving the cursor over it. Glossing was originally intended to support learner-autonomous comprehension of authentic or non-simplified texts (Davis, 1989; Holley & King, 1971; Jacobs, 1994; Lomicka, 1998). Over time, however, research interest has shifted to its potential for vocabulary acquisition. The latter is the focus of the present article.

If unfamiliar lexical items in a text are clarified to second language (L2) readers by means of glosses, then what is the likelihood that the items will be remembered beyond the reading activity itself? Owing to the numerous ways and conditions in which glossing can be implemented, giving an answer to this question will not be straightforward. Meta-analytic reviews have yielded estimates, but I will argue that these remain hard to interpret because the magnitude of the effect varies considerably from one study to the next, depending on the nature of the glosses and various other factors. It is undeniable, though, that adding glosses to texts increases the likelihood of acquisition in comparison with leaving learners without any assistance. Research has also indicated that L2 learners consider glosses helpful for vocabulary learning as well as text comprehension (e.g., Cheng & Good, 2009; Ko, 2012).

The question then is how this assistance in the form of glosses is best provided. Recurring themes in the literature include ways of presenting a gloss, stimulating cognitive engagement with a gloss, and the provision of more than one clarification for the same word (such as both a definition and a pictorial representation). In some of these areas, research findings have accumulated sufficiently for concrete recommendations to be distilled, but for several questions the answers remain provisional owing to the wide diversity in research designs and contexts.

© The Author(s), 2021. Published by Cambridge University Press

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

Before reviewing the research on the diverse ways in which glossing can be implemented, I will in the following section review the evidence that glossing per se positively influences vocabulary learning from texts, as compared with reading the same texts without the provision of glosses.

2. Estimating the effect of glossing in comparison with reading without glosses

2.1. How large is the estimated effect?

Glossing typically serves two functions: it directs the reader's attention to lexical items they may not yet be familiar with (the items are usually typographically enhanced in the text) and it clarifies their meaning (provided the reader finds the glosses clear enough). It stands to reason that glosses improve the chances of vocabulary acquisition relative to reading non-glossed texts, because in the latter case learners are less likely to attend to new words (see, e.g., Schmidt, 2001; Sharwood-Smith, 1993, on the importance of attention) and when they do try to infer the meaning of unfamiliar items, they are often unsuccessful (Bensoussan & Laufer, 1984; Laufer & Sim, 1985; Na & Nation, 1985; Nassaji, 2003). That said, it is well established that vocabulary acquisition does also happen—albeit slowly—through reading texts without the support of glosses (e.g., Day et al., 1991; Elgort & Warren, 2014; Godfroid et al., 2018; Horst et al., 1998; Pellicer-Sánchez, 2016; Pitts et al., 1989; Zahar et al., 2001). The question is thus not whether adding glosses supports vocabulary acquisition, but rather HOW MUCH MORE vocabulary acquisition can reasonably be expected thanks to glossing.

An early meta-analysis addressing this question was Abraham (2008). This analysis focused on computer-mediated glosses, that is, hyperlinked annotations that learners can access during on-screen reading. Drawing on six studies[1] that compared scores on vocabulary tests administered shortly after the reading activity, Abraham found a large average effect in favour of glossing (Hedges' $g = 1.40$). It is important to note that the effect sizes varied considerably across the individual studies, as captured by the 95% confidence intervals (0.73–1.76). In one study, Ben Salem (2006) the effect of glossing was extraordinarily large (Hedges' $g = 2.90$), reflecting a very wide gap between the scores obtained on vocabulary posttests after reading texts with versus without glosses. Such a gap does not necessarily entail a spectacular amount of vocabulary learning through glossed reading, though. It also emerges when the NON-glossed reading condition brings about no vocabulary learning, while the glossed reading does (e.g., Abraham, 2007; Ben Salem, 2006).

A possible explanation for the negligible vocabulary learning from reading non-glossed texts in these experiments is that some of the texts used were not quite comprehensible WITHOUT the glosses. When a text is not understood by learners, they will not be able to make use of context to infer the meaning of unfamiliar words (Laufer, 2020). Whether a text will be sufficiently well understood depends on numerous factors (Bernhardt, 2010; Koda, 2005), and one such factor is what proportion of the words in the text the learner already knows. The more unfamiliar words there are in a text, the more challenging text comprehension will be. When more than 5% (that is, one in 20) of the running words are unfamiliar, text comprehension will be seriously compromised (Hu & Nation, 2000; Laufer & Ravenhorst-Kalovski, 2010; Schmitt et al., 2011). Going by glossing density (e.g., 8.4% of the running words were glossed in Abraham, 2007) as a rough indication of how many words the researchers considered unfamiliar to the participants in their studies, several of the texts used in the research on glossing to date were probably not comprehensible without the glosses (see section 2.2 for more examples). If so, the question needs to be asked whether it is realistic to require learners to read a text without any assistance when the text is in fact unsuitable for this.

To find out if a non-glossed text version was sufficiently comprehensible to a given group of participants, a measure of text comprehension is useful. It so happens that the primary studies in Abraham's (2008) meta-analytic review that examined vocabulary learning also examined text comprehension, allowing Abraham to supply a list of effect sizes (p. 225) regarding how well the glossing benefited learners' performance on text comprehension measures in comparison with the non-glossed reading. These effect sizes again vary widely across individual studies, but the largest effects were observed in the studies where the non-glossed reading brought about only negligible vocabulary

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

learning (e.g., Abraham, 2007; Ben Salem, 2006).[2] Calculating the correlation between the effect sizes for vocabulary learning for the six studies and their effect sizes for text comprehension yields $r = 0.91$ ($r_s = 0.77$). In short, the POORER the comprehension of the NON-glossed texts in comparison with the glossed versions, the larger the effect of glossing on vocabulary learning tends to be.

Abraham (2008) was an early and small-scale meta-analysis. A more recent and much more extensive meta-analytic review, by Yanagisawa et al. (2020), also lends strong support to the hypothesis that the effect of glossing on L2 vocabulary acquisition is substantial. Apart from the larger scale (42 studies), Yanagisawa et al.'s meta-analysis differs from Abraham's in the following respects: (a) it includes studies on glossing in both print and computer-mediated reading conditions, (b) it includes only data sets concerning glosses that provide a single annotation per target word (see further below for multigloss and multimodal gloss implementations), and (c) besides studies that directly compare glossed and non-glossed reading, Yanagisawa et al.'s meta-analysis includes studies that compare different kinds of glosses but without a non-glossed comparison condition. The advantage of including the latter studies is that more data sets can be added to the analysis. A disadvantage is that the comparison between glossed and non-glossed reading becomes less direct, since learning outcomes from experiments where learners only read glossed texts are now being compared with outcomes from other experiments where learners read other texts (with other target words) in both glossed and non-glossed versions. Yanagisawa et al. (2020) computed the average learning gains for all the data sets regarding glossed reading and then again for all the data sets regarding non-glossed reading. They found an average learning gain from reading glossed text of 28.4% according to recall tests (e.g., where participants are presented with the L2 words and asked to supply their meaning) administered shortly after the reading activities. Recognition tests (e.g., where participants are required to match the L2 words to their correct meaning, typically using a multiple-choice (MC) format) yielded an average of 54.6%. Average gains from reading non-glossed text were much smaller: 14.1% according to recall tests and 33.8% according to recognition tests, administered shortly after the reading activities.

It is important to note that the vocabulary gains show wide variation across data sets in Yanagisawa et al. (2020), as they did in Abraham (2008). This variation is reflected by the 95% confidence intervals for the learning gains reported by Yanagisawa et al. (2020). For example, regarding the gains measured by recall tests, these CIs were 18.2%–38.7% after glossed reading and 6.4%–20.2% after non-glossed reading. Much of the variation in learning gains must be owing to differences in the design features of the studies. For example, while in most of the studies (e.g., Ko, 2012) learners read a glossed text and were given a test on the glossed vocabulary immediately afterwards in the same session, others (e.g., Jung, 2016) distributed reading tasks over separate sessions and what was labelled an immediate test by the researchers actually came with a delay regarding at least some of the reading texts. In the case of delayed tests, variation in outcomes may in part be owing to the differences in time lag between the reading activity and the test as well, ranging from one week (e.g., Watanabe, 1997) to five weeks (e.g., Nagata, 1999).

The variation in outcomes must also be influenced by the design of the test instruments, especially in the case of multiple-choice tests (MCTs). For example, some MCTs used distractors that resembled the meaning of the target words (e.g., Miyasako, 2002), making the test more challenging than using semantically distinct options (e.g., Yoshii, 2006). Sometimes the distractors in the MCT were meanings that the participants had encountered during the reading activity (e.g., Jung, 2016), while at other times participants could eliminate proposed meanings because they had not met them at all during the reading activity (e.g., Zhao & Ren, 2019). Furthermore, the fewer the number of response options in MCTs, the greater the likelihood of lucky guesses and thus the risk of inflated success rates. In the collection of primary studies examined by Yanagisawa et al. (2020), the number of options in the MCTs range from three (e.g., Ko, 1995) to eight (e.g., Shahrokni, 2009).

All this variation makes estimating the overall effectiveness of glossing complicated, especially when data sets come from studies with and without direct comparisons of glossed and non-glossed reading. Another factor worth mentioning is that, in a few of the primary studies included in Yanagisawa et al. (2020), some or all of the target words occurred several times in the texts (e.g., Hulstijn et al., 1996; Nagata, 1999; Rott et al., 2002). Repeated encounters with words help their acquisition (Uchihara et al.,

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

2019, for a meta-analysis), and this holds true also for glossed texts (e.g., Teng, 2020). In a comparison of data sets from glossed or non-glossed reading conditions, an uneven distribution of datasets from experiments where the target words were encountered several times by the participants is thus also likely to influence the outcome.

2.2. Matters of text comprehensibility

Returning to an observation made previously, it is striking how little vocabulary learning happened through NON-glossed reading in several experiments (e.g., Jung, 2016; Samian et al., 2016; Watanabe, 1997; Zhao & Ren, 2019). It needs to be reiterated that, in some studies, the non-glossed reading condition was not altogether realistic. If it is deemed necessary to add a gloss per every 20 or so running words in order to make the text comprehensible for a given learner group (e.g., Arpaci, 2016; Azari et al., 2012; Jacobs et al., 1994; Sadeghi & Ahmadi, 2012), then the flipside of the coin is that, WITHOUT the glosses, the text is probably not going to be comprehensible. A well-informed teacher would probably not choose such material for their students' unassisted reading in the first place. It is true that highly motivated learners may try to read texts that are far beyond their proficiency level in their spare time, for example because the texts concern a subject of personal interest, but such learners will then probably consult resources such as (online) dictionaries when they fail to comprehend a passage that they are really interested in. The participants in the non-glossed reading conditions in the experiments reviewed here had no access to any resources outside the actual texts.

To be fair, several of the research reports (e.g., Cheng & Good, 2009; Ko, 2012; Zhao & Ren, 2019) do include good justifications for the choice of texts (along with justifications for the choice of target words), with mention of piloting efforts and consultation of the students' instructors. Often, though, the sole assurance given that the texts were likely to be comprehensible is a readability measure such as the Flesch-Kincaid formula (e.g., Farvardin & Biria, 2012; Samian et al., 2016; Shahrokni, 2009; Watanabe, 1997). It is worth noting that this formula concerns mostly syntactic complexity and was originally intended for estimating readability in a first language (L1). It may not suffice as assurance about text comprehensibility for L2 readers, for whom the lexical profile of a text is likely to matter more (Qian & Lin, 2020, pp. 68–69). Administering a Vocabulary Levels Test (Schmitt et al., 2001; Webb et al., 2017) to estimate whether learners will be able to cope with a given text is probably a better assurance. There are also several research reports where no assurance is given regarding the suitability of the chosen texts for the given learner groups (e.g., Abraham, 2007; Lee et al., 2016; Moradan & Vafaei, 2016; Nagata, 1999; Varol & Erçetin, 2016).

In addition, some researchers deliberately chose reading texts about a topic that was novel to their participants (e.g., Arpaci, 2016; Jung, 2016). This was done to control for topic familiarity, because topic familiarity is one of the factors known to aid text comprehension and, in turn, the ability to make effective use of context to infer word meanings (Pulido, 2004). Without background knowledge about the subject of a text, inferring the meaning of unknown words is bound to be especially challenging, and so vocabulary gains from non-glossed reading will likely be particularly small. Moreover, some researchers went to great lengths to make sure that the meaning of the target words was not inferable from context (Ben Salem, 2006), thus making vocabulary learning from the non-glossed version virtually impossible. Using texts that are selected, modified or created with a view to preventing successful inferencing can be helpful for the researcher to isolate the effect of glosses, because any learning of new word meanings from reading the glossed version of such a text will then be attributable to the glosses (rather than successful inferencing). At the same time, one may wonder whether teachers often ask their students to (a) read texts on totally unfamiliar topics without any preparation and (b) work with texts that have been specially modified to prevent successful inferencing.

There is no doubt that glossing assists vocabulary learning, but, as I argued above, the gap between the amount of vocabulary learning from glossed and non-glossed versions of texts could be narrower in actual pedagogic practice, where it is hoped that learners read texts that are within their comfort zone, and where consequently also non-glossed versions provide the learners with a chance of picking up at least a few new lexical items.

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

2.3. Diversity in reading tasks

What the participants were asked to DO with the reading materials is bound to influence learning outcomes as well. In only a small number of the studies were students simply asked to read texts for comprehension. In most, the students were given tasks that engaged them intensively with the texts and their glosses. For example, students were asked to read a text with a view to reproducing its contents (e.g., Gan, 2014; Ko, 2012; Rott et al., 2002) or with a view to answering numerous questions that prompted the students to inspect the glosses (e.g., Samian et al., 2016). The effectiveness for vocabulary learning of text comprehension tasks that direct learners' attention to target words has been demonstrated convincingly by Peters et al. (2009). Sometimes, the participants were explicitly told they should examine the glosses (e.g., Aizawa, 1998; Lee et al., 2016; Rassaei, 2018; Salehi & Naserieh, 2013). Still, other studies examined the merits of MC glosses, which require the reader to choose which of two or more proposed word meanings matches the contextual use of the target word (see section 3.4). In some experiments (e.g., Rott, 2005), the students were asked to first read a text without glosses and then with glosses, and to think aloud while interacting with the text and the glosses (also see section 4).

In most of the studies, learners worked at length with a short text passage (sometimes just one short paragraph), while in other cases they read several pages at a more natural pace. For example, in Arpaci (2016) students spent 25 minutes on a 250-word passage, whereas in Hulstijn et al. (1996) students spent the same amount of time on a text of over 1,300 words. It is likely that learners examine language features (including glossed lexical items) more closely if they are asked to work at length with a short text than when they read a long text at a more natural pace. It is unfortunate that few authors mention both text length and how long the reading activities took. Among the few studies included in Yanagisawa et al.'s meta-analysis that do mention this information, the time-on-text ratio varies widely, from just seven words per minute (Nagata, 1999) to 92 words per minute (Zhao & Ren, 2019).

To examine whether the time-on-text ratio is likely to influence vocabulary learning gains—as one would expect—I ran the following correlational analysis of a handful of data sets[3] included in Yanagisawa et al. (2020). For the sake of comparability, these data sets are exclusively from studies that used: (a) texts in which the words of interest occurred once, (b) meaning-recall posttests in the same session as the reading activity, and (c) marginal glosses offering a single meaning (no MC glosses—see section 3.4). I found a negative correlation between the average number of words dealt with per minute by participants and their vocabulary gains: $r = -0.412$ ($n = 6$). So, it is indeed when learners spend a long time working with short texts that substantial vocabulary gains tend to occur. This helps to put the average learning gains that Yanagisawa et al. (2020) found into perspective, because a large number of primary studies included in their meta-analysis engaged learners with short texts for considerable lengths of time, whereby these learners possibly considered these texts as materials for language-focused (and vocabulary-focused) study rather than materials for 'mere' reading practice.

Still, articles reporting research on glossing almost invariably characterize the observed learning as 'incidental', that is, as acquisition from activities that are focused on communicative content rather than the language code. It is true that participants in most of these studies were neither instructed explicitly to memorize the glossed items, nor were they informed a test about them would follow. Absence of test announcement is now considered by many scholars as the most straightforward way of operationalizing the distinction between incidental and intentional learning (e.g., Hulstijn, 2003; Webb, 2020, for discussion). That said, if students are given a decontextualized text passage that is only about 50 words long and in which no fewer than five (10%) of the words are highlighted and clarified in a glossary (Azari et al., 2012), then is it not likely that the students will experience this input as material for deliberate vocabulary study?[4] This is an extreme example, but, as mentioned previously, there are numerous studies that used short texts where almost every sentence had a glossed word (Abraham, 2007; Cheng & Good, 2009; Nagata, 1999; Yoshii, 2013). Even though they are

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

not explicitly forewarned about a vocabulary test, might the nature of such reading conditions not create an expectation that such a test will follow—especially if the participants were previously given a pretest about the words that they subsequently see clarified in a glossed text?

This discussion of whether vocabulary learning from reading glossed texts should always be considered incidental is arguably just of academic interest, while practitioners are concerned first and foremost with what is best for their students' learning outcomes, regardless of the cognitive mechanisms presumed to generate those outcomes. That said, if reading a glossed text is perceived by learners as a vocabulary study activity, then a practical side effect of this may be that the text itself will no longer be of much interest to them. There are indeed anecdotes in publications that participants spent more time studying the glosses than reading the text (e.g., Abraham, 2007, p. 105). In some of the experiments, measures of text comprehension or recall of text content showed either no effect of glossing (e.g., Cheng & Good, 2009; Jacobs et al., 1994; Jung, 2016) or in fact better results for the NON-glossed condition (e.g., Plass et al., 2003). In such cases, the glosses clearly did not serve the purpose that they were originally intended for—scaffolding text comprehension.

It is probably when learners are given a relatively long text with moderate use of glossing (e.g., glosses for under 2.5% of the running words, as was done in Hulstijn et al., 1996)[5], and told to read for text comprehension that vocabulary gains from glossed reading can safely be attributed to incidental acquisition—where incidental acquisition is characterized not only by the absence of test announcement but also by the content-focused nature of the activity. In Hulstijn et al. (1996), the average learning gain (for words that occurred once in the text) from reading the glossed version of the text was 16% according to a meaning-recall test. This was better than the gain from reading the non-glossed version (just 4%), but clearly less impressive than the average gain from glossed reading calculated in Yanagisawa et al.'s (2020) meta-analysis (28.4%). Apart from exceptions such as Hulstijn et al. (1996), reading activities in studies on the effects of glossing have overwhelmingly used short texts with high glossing densities. It would benefit future meta-analyses—especially ones interested in incidental acquisition—if more studies with longer texts and moderate use of glossing were conducted. That said, practitioners may justifiably ask the question why glossing should not also be tried as an explicit and intentional language-focused intervention, anyhow (e.g., Khezrlou et al., 2017). For example, how effective is reading glossed text in comparison with reading non-glossed text when the learners are told beforehand to expect questions about text content AND vocabulary?

So far, I have presented evidence from meta-analytic reviews that glossing positively influences vocabulary learning, but have also argued that the general magnitude of this positive effect needs to be interpreted in a nuanced and cautious way, considering the wide diversity in study designs and instruments used in the collection of primary studies. The discussion so far concerned a broad comparison of glossed and non-glossed reading. Absent from the discussion so far has been the question whether some implementations or types of glossing are more effective than others. So, let us now turn to such comparisons. Topics to be discussed include the location of glosses (e.g., in the margin or elsewhere), the language of glosses (L1 or L2), the modality of glosses (e.g., a combination of text and picture), ways of stimulating cognitive engagement with glosses (e.g., presenting the glosses as MC quizzes), and the use of glosses when target words occur several times in the same text.

3. Evaluating ways of implementing glosses

3.1. The location of glosses

One question that has attracted a fair amount of interest is where a gloss should be provided for learners to make the best use of it. Research findings suggest that listing the words and their meanings at the very end of a text, in the form of a glossary, is not optimal (e.g., Cheng & Good, 2009; Jacobs et al., 1994), probably because learners are reluctant to interrupt the flow of reading to look for an unfamiliar word in the glossary (Ko, 2012). Instead, placing the gloss closer to the location of the unfamiliar word

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

in the text, so it can be inspected with less interruption, appears the better choice. This is often done by providing the gloss in the margin next to the line where the unfamiliar word occurs. Another possibility is to provide it within the text, directly adjacent to the unfamiliar word. However, according to Yanagisawa et al.'s (2020) meta-analysis, which examined the location of glosses as one of several moderating variables, such in-text glosses appear comparatively INeffective for vocabulary uptake. This was based on data from only four studies that used in-text glosses, however, and it may be skewed by one data set (Watanabe, 1997), which implemented in-text glossing as the provision of so-called appositives, inserting a clarification after the new word without overt indication that it serves this clarification purpose (as I did just now regarding the term *appositive*). When some of the participants were asked to infer the meaning of the target words with the text at their disposal, they often failed to do so, which suggests they simply did not recognize the purpose of the appositives. In other studies, where in-text glosses were more clearly signalled as word clarifications, these did yield comparatively good results (Ouyang et al., 2020; Zhang & Webb, 2019, for recent examples). A re-assessment of the effectiveness of in-text glosses may become feasible when more studies comparing in-text and marginal glosses become available.

The number of glossed words may be a moderating variable in this specific evaluation as well (Nation, 2013, p. 242), because, if a high number of words are glossed in the margin of the text, then this will clutter the margin and it will become difficult to position each gloss next to the text line where its corresponding target word occurs—which may possibly cause cross-item confusion. Even if researchers choose a text in which 'only' 5% (i.e., one in 20) of the running words are unfamiliar to a given group of readers, providing a marginal gloss with each occurrence of an unfamiliar word would amount to a gloss next to almost every line of a text such as the one you are reading, and it is likely that occasionally a single line will include two unfamiliar words. If so, it may be more judicious to use a glossary at the end of the text after all. Alternatively, one could select only some of the unfamiliar words for marginal glossing and hope that the meaning of the remaining unfamiliar words will become better guessable. Selecting 2% would amount to about three marginal glosses in the case of the present paragraph, for example, making it easier to put the glosses next to the relevant text lines and making the amount of information presented in the margin better manageable for the reader. That said, more research, including studies of learners' perspectives, is needed to support this recommendation.

A fair number of studies about glossing (e.g., Lee et al., 2016; Yeung, 1999) have used a 'cognitive load' perspective (Sweller, 1994) to evaluate how excessive glossing may interfere with learners' ability to process a text and the accompanying information. In the case of computer-mediated reading, the above problem of cluttering text margins with glosses can be avoided, as the glosses can be made accessible one at a time through hyperlinks. This may also hold an advantage over glossaries added at the end of the text, because the reader just needs to mouse-click on the word or move the cursor over it to access the gloss, which is less cumbersome than scrolling down to the end of the text to find the relevant information (especially if the text is long). That said, where the hyperlinked glosses are presented probably matters as well. If the hyperlinked gloss is presented on a separate screen (e.g., Rassaei, 2018), then readers are temporarily taken out of the reading text and will need to re-connect with it. Learners may therefore prefer a design where a pop-up window appears on the same screen as the text passage (e.g., Bowles, 2004; Lee et al., 2016). A recent study by Varol and Erçetin (2019) indeed suggests that learners are more likely to access hyperlinked glosses if these appear in a pop-up window on the same screen as the text than if they appear on a separate screen. This preference is likely to be particularly strong if there are many hyperlinks and text comprehension is challenging (as was the case in Varol and Erçetin's experiment). On the other hand, if there are just a few hyperlinks and the learners find the text manageable, then they may not mind studying the glosses on a separate screen. Considering the evidence in support of retrieval-from-memory practice (e.g., Van den Broek et al., 2018), there may even be grounds for hypothesizing that briefly holding the information from the gloss in memory and then reconnecting it to the instance of the word in the actual text is more beneficial for retention than seeing the gloss together with the word in the text (see also section 3.5).

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

3.2. L1 versus L2 glosses

The meta-analysis by Yanagisawa et al. (2020) examined the language of glosses (L1 translations versus L2 definitions) as a moderator variable and found an advantage for L1 glosses. This was confirmed by yet another recent meta-analytic review (Kim et al., 2020), which compared the benefits of L1 and L2 glosses by specifically examining a collection of primary studies in which language choice was the independent variable of interest. In this collection, 21 studies included measures of vocabulary learning. L1 glossing was found to be the most beneficial, although the overall effect sizes in its favour were on the small side (g = 0.44 and 0.28 in immediate and delayed posttests, respectively). Interestingly, the advantage of L1 glosses was most pronounced in the case of studies with learners at low proficiency levels, likely because it is low-proficiency learners who find L1 glosses much easier to understand than L2 glosses.[6] It stands to reason that comprehension can only be supported well by glosses if the glosses themselves are comprehensible to the learner, meaning L2 glosses should be carefully designed to match the learners' proficiency, for example by avoiding words that the learners might not yet be familiar with.

It is surprising in this regard, that information about the L2 glosses that were used is absent in most of the research reports (e.g., Arpaci, 2016; Ertürk, 2016; Farvardin & Biria, 2012; Jacobs et al., 1994; Rouhi & Mohebbi, 2012; Shiki, 2008; Taheri & Zade, 2014; Zarei & Hasani, 2011) while others provide just one or a couple of examples (e.g., Mitarai & Aizawa, 1999; Miyasako, 2002; Salehi & Naserieh, 2013). This makes it impossible to evaluate how clear to the study participants these L2 definitions were in comparison with the L1 glosses. Even under the assumption that they were crystal clear, clarifications in L2 may still take more time and effort to process simply because they are longer than L1 translation equivalents, especially when abstract words need to be defined. For example, in Salimi and Mirian (2019) the L2 definitions were up to 14 words in length and in Öztürk and Yorgancı (2017) many of the definitions were over 20 words.[7] It is no wonder in such cases that learners are reluctant to read L2 glosses and by far prefer L1 glosses (e.g., Salimi & Mirian, 2019). However, this preference may change when the L2 glosses are concise and clear enough (e.g., Ko, 2012). Still, even if the L2 glosses are kept concise, readers tend to process L1 glosses faster (e.g., Kang et al., 2020).

If learners (especially at low proficiency levels) are less inclined to inspect L2 glosses than L1 glosses for the above reasons, then this may hold true in particular when the reading activity is done under time pressure and when an excessive number of glosses is available.[8] It is indeed striking that the studies that showed poor effectiveness of L2 glosses compared with L1 glosses tend to be ones that used reading materials with a high glossing density (e.g., Arpaci, 2016; Yoshii, 2006).

Eye-tracking research would be welcome to explore the potential interaction between gloss language and reading behaviour at different proficiency levels, at varying glossing densities, and in timed versus untimed reading conditions. If such eye-movement research were to confirm that L2 glosses are more often ignored than L1 glosses, then the question should be asked what could be done to increase the likelihood that learners DO make good use of L2 glosses. This question is relevant because many textbooks and other instructional materials for language learning are produced for a global market or are meant for groups of learners with mixed L1 backgrounds. In those cases, L1 glosses are not a viable option (although this may change as computer-mediated reading becomes increasingly common and learners may be offered options to customize their reading experience).

3.3. The addition of pictures

Word meanings can also be elucidated by pictures. Several studies have compared the effectiveness of text-only, picture-only, and text-plus-picture glosses (e.g., Kost et al., 1999; Yoshii & Flaitz, 2002) or just text-only and text-plus-picture glosses (e.g., Boers, Warren et al., 2017). Collectively, the results suggest that text-plus-picture glosses are the most beneficial. This, at least, is the conclusion reached by a meta-analytic review by Yun (2011), although the average weighted effect was on the small side (Hedges g = 0.37).[9] This analysis was based on only ten studies at the time, however, and there are additional studies that did NOT furnish evidence in favour of the addition of pictures (Acha, 2009; Ben Salem, 2006; Boers, Warren et al., 2017).

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

Half of the studies included in Yun (2011) were computer-mediated reading experiments—many of which were inspired by Mayer's (2009) theory of multimedia learning—where learners were invited to click on highlighted words to call up one or more glosses (Akbulut, 2007; Chun & Plass, 1996; Jones & Plass, 2002; Plass et al., 1998, 2003). In some reading conditions, only one gloss (a textual clarification OR a picture) was available for the target word, while in another condition the learners could call up two separate glosses (a textual clarification AND a picture). The two-gloss condition was found to be the most beneficial for word learning. Although authors of these studies tend to ascribe this effect to multimodality (i.e., exposure to both a textual and a pictorial clarification), an alternative explanation is that the availability of two separate glosses prompts learners to check information about the words twice instead of just once (Plass et al., 1998). A meta-analysis by Ramezanali et al. (2021) compared reading conditions with one and two hyperlinked glosses at the readers' disposal and indeed found a general advantage of having two glosses to inspect.[10] This advantage is not necessarily owing to multimodality per se, however. After all, looking up a word meaning more than once improves the likelihood of word learning also if it concerns only a textual clarification (e.g., Montero Perez et al., 2018; Peters et al., 2009; Rott, 2007). To separate the hypothesized effect of multimodality from that of multiple lookups, future experiments could offer two hyperlinked glosses per target word where in some reading conditions participants are informed that both glosses are textual or both are pictorial, while in another condition they are told one gloss is textual and the other pictorial. Computer log data could reveal under which condition the participants look up a word more than once and posttest data could then reveal if the combination of text and picture leads to better recall when the number of lookups is statistically controlled.

Turning back to Yun's (2011) meta-analysis, this also included five studies where the multimodal information (text and picture) was provided IN A SINGLE GLOSS. Three of these found an advantage of text-plus-picture glosses. In Kost et al. (1999), these glosses led to better performance than only textual or only pictorial glosses in one of three posttests. Similar results were obtained by Yoshii and Flaitz (2002) and Yoshii (2006). A potential issue with the latter two experiments (which used the same materials) is that the text-only and the picture-only annotations appear to have underspecified the meaning of target words. The examples of gloss types given in Yoshii and Flaitz (2002, p. 37) concern the verb *dash*. This was defined as 'to move very quickly' in the textual gloss and elucidated by a drawing of a running figure in the pictorial gloss. In a posttest where the learners were required to supply the meaning of *dash*, two points were awarded if this included both the semantic features 'fast' and 'run' and one point was awarded if only 'run' was included (p. 53). Since the manner of motion (run) was not specified in the textual gloss, it is not surprising that learners in that reading condition failed to obtain good scores. It was the group of participants who had seen the combination of the textual and the pictorial information who stood the best chance on the test, because the two pieces of information TOGETHER provided the meaning that was needed to obtain full points in the test. If the textual annotation had been more precise, such as '*dash* = to *run* somewhere quickly', the results might have been quite different. Yeh and Wang (2003) is another article claiming superiority of text-plus-picture glosses over L2 definitions only, but in this study the posttest scores were in fact not significantly different. The final study in Yun's (2011) meta-analysis, Al-Seghayer (2001) used a within-participant design where seven words in a text were glossed by L2 definitions, seven others by a definition plus a picture, and another seven by a definition plus a video clip. The latter gloss type—which very likely invites longer engagement with the gloss—led to the best outcome, but no significant difference was found between definition-only and definition-plus-picture glosses.

In sum, when text-plus-picture glossing was found to be more effective in some of the experiments, it is not always clear WHY it was more effective, because other factors than multimodality may have played a part. Many authors invoke *Dual Coding Theory* (Paivio, 1986) to explain the proclaimed benefits of adding visuals to glosses. According to this theory, words that are associated with a mental image of their referents are easier to remember than words that only have a propositional representation in the mental lexicon. However, it is not entirely clear how this theory helps to explain the positive findings about pictorial glosses, because the words glossed with pictures usually have concrete

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

meanings and are thus mentally imageable anyhow. If you learn that the French word *cheval* means 'horse', then surely you can conjure up a mental picture of the animal without being shown a picture. It is probably when concrete referents that you are not yet familiar with are introduced, such as objects or actions that are technical or that are alien to your own culture, that visuals are the most helpful (Mayer & Anderson, 1992; Mayer & Gallini, 1990). If Dual Coding Theory does not offer a convincing explanation for the observed benefits of multimodal glosses, then an alternative explanation may be that pictures attract attention where verbal definitions alone fail to do so. An eye-tracking study by Warren et al. (2018) indeed revealed that marginal glosses providing only a definition (in L2) in their experiment were ignored by participants far more often than glosses containing a picture as well as a definition.

3.4. Stimulating cognitive engagement

It is possible that glosses fail to leave robust mental representations of word meanings when learners 'passively' take in the information and then carry on reading without giving further thought to the glossed item—which is a likely scenario when the reader is (naturally) interested first and foremost in the content of the text rather than how it is worded. Hulstijn (1992) explored using MC glosses for the purpose of promoting more cognitive engagement with glossed words. MC glosses require readers to closely examine the word's use in the text to determine which of the proposed meanings in the gloss best matches it. Unless the correct choice is immediately obvious (which arguably defeats the purpose of triggering an interpretation effort), a downside of MC glosses is that learners may pick the wrong meaning. For example, Watanabe (1997) tried using glosses where two meanings were proposed, but these two options (e.g., *cribs = (1) babies' beds; (2) small beds*) were often both plausible matches for the reading context. Since there was no further evidence in the text for students to confirm or revise their interpretations, it is not surprising that the posttest scores of students who received MC glosses did not trump those of students who had been provided with standard, single-meaning glosses. An alternative is to give immediate feedback on the choices made in the MC glosses. In Nagata (1999), students were explicitly asked to make a choice between two translations, after which they were informed if their choice was correct. This implementation of MC glossing led to better meaning recall than standard glossing. In Nagata's (1999) experiment, the incorrect translations were not plausible, and so the risk of error was minimal. Yoshii (2013) conducted a similar study but with a greater likelihood of wrong choices. Yoshii's (2013) article is exceptional in that it includes helpful data per student regarding which items they got wrong in the MC glosses and in the MC posttests. Using these data, I tallied how often a wrong choice during reading was followed by a wrong choice in the test, despite the corrective feedback given at the reading stage. Of the incorrect responses to the MC glosses, 22.5% and 35% were followed by incorrect responses in the immediate and the one-week delayed posttest, respectively. This is appreciably higher than the average error rates in these posttests regardless of whether the MC glosses were answered correctly, which were only 14% and 25.5%, respectively. Evidently, the effectiveness of MC glosses can be compromised when learners make wrong choices at the initial learning stage, even if corrective feedback is given (e.g., Boers, Dang et al., 2017; Strong & Boers, 2019, for more evidence of this kind). The available collection of studies examining the merits of MC glosses again illustrates that different implementations of a given technique can yield markedly different learning outcomes,[11] and so it also illustrates the need for caution and nuance if one wishes to recommend the use of MC glosses. It will be helpful if, like Yoshii (2013), future studies on MC glosses include data on participants' response behaviour as an indication of how challenging the MC glosses were.

There may be additional ways of inviting cognitive engagement with glosses. One possibility regards polysemy. In Boers (2000), for example, students majoring in commerce and economics read a text containing lexical items such as *hurdles* (in the sense of difficulties to be overcome), *wean off* (as in *weaning an industry off state support*) and *weed out* (as in *weeding out corruption*). Although these items occurred in the text in these abstract meanings, they also have literal meanings,

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

from which the abstract uses are derived through metaphorization: *hurdles* are the barriers that athletes jump over in track sports such as steeplechase, *weaning off* means gradually stop breastfeeding one's baby, and *weeding* refers to pulling out unwanted plants in one's garden. Half of the students in the experiment read the text with marginal glosses that only gave these literal meanings; the other half received glosses paraphrasing the abstract meanings in the text (e.g., *hurdles* = difficulties; *weaning off* = making independent; *weeding out* = removing). The students who had received only the literal meanings obtained the better scores on a productive recall test three days after the reading activity, even though the target items were again embedded in abstract contexts. Presumably these students had needed to invest some effort into working out the connection between the literal meaning given in the glosses and the contextual use of the items. This is perhaps where Dual Coding Theory does offer a plausible account—linking abstract word uses to their literal origins increases their imageability and can thus make them easier to remember. If glosses that prompt learners to think about the connection between related meanings of a lexical item can make the item more memorable, then this is promising because polysemy is very common. Boers (2000) was just one experiment, however, and one with serious shortcomings (such as lack of pretesting). Approximate and conceptual replications would therefore be welcome.

In essence, the purpose of MC glosses or other glosses that do not present readers immediately with the precise contextual meaning of lexical items is to create temporary ambiguity, where readers are made to wonder for a little while what exactly the contextual meaning of the item is. It does appear important, however, to reduce the risk of wrong interpretations lingering in memory. Encounters with the same lexical item in additional contexts should be helpful.

3.5. Increasing the number of encounters

Some of the studies on glossing have used texts with several instances of the same target words. For example, Rott (2007) inserted four instances of the same unfamiliar words in short texts (of about 350 words in length) and found that providing a gloss on each instance had a greater positive effect than presenting the gloss only with the first instance. It is likely that seeing the same words glossed several times on the same page left the students with the impression that these must be very important and thus worth studying. However, with target words occurring more than once in a text, it also becomes possible to create reading conditions where learners need to recall information from a previously seen gloss that they no longer have access to. Huang and Lin (2014) incorporated target words three times in a text, but each occurrence of the target word was displayed on a new screen, and the students could not return to previous screens.[12] In one reading condition, the words were glossed on every occurrence, but in another reading condition only the first and third instance were glossed. The students in this condition were asked to supply the meaning of the words when they were met the second time. There was a third reading condition, where the students saw the words first without glosses and were asked to infer their meanings before seeing the glosses on the following pages. The best posttest results were obtained when students had seen the words explained first, then needed to retrieve this information from memory on the second encounter, and then saw the gloss again.

Huang and Lin's (2014) study adds to an already large body of evidence of the effectiveness of retrieval of information from memory (e.g., Barcroft, 2007, 2015). Posttest performance was not as good when students had first been required to infer word meanings before seeing the glosses, and this was found to be owing to the instances where students had started off with incorrect inferences. That incorrect guesses or inferences can interfere with later recall of correct form-meaning correspondences has been observed in a few other studies, albeit not ones directly concerned with glossing (e.g., Elgort, 2017; Elgort et al., 2020; Mondria, 2003). The students who engaged in neither inferencing nor retrieval in Huang and Lin (2014) because the target words were always accompanied by their glosses obtained the poorest posttests scores in this experiment. Note that, unlike in the other two reading conditions, the students in this condition were not explicitly prompted at any time to think about the meaning of the target words. Approximate replications in which word-focused tasks are present

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

or absent in all the reading conditions would be welcome (see my comment in section 2.3 about incidental vs. intentional learning).

A combination of MC glosses and repeated encounters may also be a fruitful way of first arousing curiosity about the meaning of an item and then offering further opportunities to fine tune one's interpretation. Rott (2005) used a text in which target words occurred four times each and were glossed only on their first instance. The participants were asked to think aloud as they read the text. The glosses were either MC glosses or standard, single-meaning ones. The first instance of each of the target words in the text did not come with much supporting context to pinpoint their meanings, but the subsequent instances were accompanied by clearer clues so as to guide learners to the correct form-meaning correspondence. The MC glossing condition was found to lead to better word retention, because—as evidenced by the think-aloud data—it aroused curiosity about the correct meaning of the target words and, when students initially made a wrong inference, the subsequent instances of the target word helped them to rectify this (an opportunity for rectification that was missing from other studies on MC glosses that I reviewed in section 3.4, such as Yoshii, 2013). Such findings suggest that temporary ambiguity can trigger a quest for word meaning that can be quite beneficial—provided this quest is successful.

Temporary ambiguity may also have played a part in a study by Warren et al. (2018). Target pseudo-words occurred three times in a story presented on a series of computer screens and were glossed in the margin of the text on their first instance only. The other two instances of each target word occurred on subsequent screens and the participants could not return to the earlier screens to revisit the glosses. Three kinds of glosses were used, depending on the reading condition that the participants were assigned to: a textual clarification in L2, a picture, or a combination of text plus picture. The learners' eye-movements were tracked as they read the text. The text-only glosses were the least effective, but this was in part owing to the lack of attention they attracted—they were ignored about 20% of the time, while glosses containing a picture were seldom ignored. The expectation was that the richest representation of meaning (i.e., the text-plus-picture glosses) would have the greatest positive effect on retention, but instead it was the picture-only glosses that emerged as the most beneficial. The authors speculated that a picture alone leaves some uncertainty about the precise meaning of a word, which then influences readers' engagement with later contextual instances of the same word. For example, one of the target words denoted a sleeveless jacket. A picture of a sleeveless jacket as such does not stipulate that absence of sleeves must be a defining feature of the word. It was by processing the subsequent text passages that readers could ascertain this, for example, when the protagonist of the story, who was wearing a sleeveless jacket, was said to have scratches on his arms after clambering through thorny bushes.

So, what predicts learning may not only be the extent to which learners engage with a gloss, but also how the gloss influences engagement with the remainder of the text and with the target words when these are encountered again.

4. Establishing cause-effect relations: To what extent do learners engage with glosses?

It is hard to determine why one implementation of glossing has a greater effect than another in the absence of data about whether learners engage with the available glosses in the first place and, if they do, how they engage with the glosses. The same applies to the instances of the target words in the text. Some researchers have used retrospective questionnaires to collect information on how learner-participants experience glossed texts (e.g., Ko, 2012; Salehi & Naserieh, 2013), but the limitation of this is that participants may not recall precisely what they did or were thinking when they performed the reading task.

In the case of hyperlinked glosses, computer logs can provide data on whether and how often students looked up a word and whether this influenced later recall. While this does not as such reveal precisely why a word was looked up and how well the learner processed the information, it at least helps to ascertain that the word and its gloss attracted attention. If the target word occurs more than once in the text but has a hyperlink only on its first occurrence, then the computer logs will

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

unfortunately not capture if the learner also pays special attention to the instances of the word where there was no hyperlink.

In comparison, eye-tracking (e.g., Ouyang et al., 2020) can provide not only data regarding attention to the glossed items but also regarding additional instances (if any) of the target words in the text. Moreover, it can capture what part of a gloss attracts attention the most (which may be useful, for example, to evaluate the respective contributions of pictorial and textual information in a multimodal gloss) and it can capture how a reader's eyes move back and forth between a gloss and the instance of the word in the text (which may be indicative of the effort the reader puts into establishing a correspondence between the information in the gloss and the contextual use of the word).

All the same, eye-tracking has its limitations, too. While eye-tracking data reveal whether something was attended to, how often and for how long, what precisely went on in the reader's mind remains elusive. For example, in the aforementioned study by Warren et al. (2018), which found comparatively good vocabulary acquisition from reading a story with only pictorial glosses, the eye-tracking data were compatible with the account that the pictorial glosses left some vagueness about the precise meaning of the target words, which then prompted the readers to more closely examine the subsequent occurrences of the word. However, in the absence of data to confirm that this was indeed what the participants were experiencing, this account must remain speculative at best.

A way of eliciting information about what goes on in readers' minds as they process glossed texts is to ask them to think aloud while reading (e.g., Bowles, 2004; Rott, 2005; Yanguas, 2009). A downside of this is that the requirement to think aloud while reading likely influences how readers engage with the text because it heightens their meta-cognitive and meta-linguistic awareness, and so the findings are not necessarily generalizable to natural, silent reading. In an experiment by Godfroid and Spino (2015), for example, participants who were asked to think aloud while reading obtained higher text comprehension scores than participants who read the text silently.

An alternative is using stimulated recall interviews, where participants are presented shortly after the reading activity with behavioural data such as their recorded fixations on words/glosses or their recorded mouse-clicks on hyperlinks, and where these data serve as prompts for the participants to recollect what they were thinking at the time. This can reveal aspects of the reading experience that the participant has episodic memories of (provided the participant's verbal report of those memories is reliable), but, unlike eye-tracking, it does not reveal the attentional processes that go on below the level of consciousness (see, e.g., Godfroid et al., 2013, for discussion).

It is therefore a combination of 'online' and 'offline' data that will help to determine cause-effect relationships best (e.g., Godfroid & Schmidtke, 2013). An excellent example of such data triangulation in the context of glossing research is a study by Jung and Révész (2018), which examined task effects on reading behaviour. Learners in this experiment were presented with jumbled text passages and asked to put these in the correct order. The texts were accompanied by L1 glosses in the margin. Eye-movement data revealed that these glosses attracted relatively little attention. Stimulated recall data helped to explain this: the glossed words were not experienced by the participants as essential for completing the task. Instead, the nature of the task had oriented their attention to discourse connectors and other hints that helped to put the passages in the right order. This study thus also demonstrates the point I made earlier (in section 2.3) that what learners are asked to do with reading material will influence what use they will make of available glosses.[13]

5. Glossing in practice

Publications about glossing almost invariably include an assertion in their introductory section that glossing is very common in L2 learning contexts, and that this is one of the motives for examining its effectiveness. Seldom, however, is this assertion substantiated, and, when sources are cited, these are typically previous research articles that made the same assertion. One exception is Davis (1989), but this study regards annotations to support comprehension of literary texts such as excerpts of classic literary works compiled for students in higher education. These annotations concern not only

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

lexical items but various points of philological interest, including cultural and historical references, and are thus rather different from the use of glosses in the research strand discussed here.

Are vocabulary glosses common in a lot of contemporary L2 course materials? I browsed through recent editions of six internationally marketed English as a foreign language (EFL) textbooks[14] for intermediate proficiency level and found regular use of glossing in only one of them, *New American Inside Out*, which provides glossaries below text passages. These glossaries sometimes concern an appreciable number of lexical items, sometimes close to 3% of the running words in the text, but this is nonetheless moderate in comparison with the copious number of glosses in some of the experiments discussed earlier (e.g., 7% in Cheng & Good, 2009; Nagata, 1999). The textbook *Straightforward* also occasionally adds word clarifications below a text but does this very sparingly. In another textbook, *Interchange*, texts are sometimes followed by a question where students need to choose which of two or more proposed meanings matches that of a word in the text, which resembles a MC gloss. These observations are about the so-called student books of these textbook series, that is, the books that are used in the classroom. Perhaps textbook writers consider glosses unnecessary in such books because a teacher will be available to clarify unfamiliar words, anyhow. I therefore also browsed through the companion workbooks, which tend to serve the purpose of independent practice (or homework) but found no glossed texts in them.

It is of course very well possible that textbooks produced for other target languages or produced for local EFL contexts make much more use of glosses, especially L1 glosses—which, according to Kim et al. (2020) and Yanagisawa et al. (2020), tend to be more effective than L2 glosses (see section 3.2). It is also possible that glossing is more common in supplementary reading resources for language learners, such as EFL magazines (Cheng & Good, 2009). Language-for-specific-purpose courses might also use glosses more often, to clarify or illustrate technical or specialized vocabulary—although it seems equally likely that such terms are defined in the body of a text because they are an integral part of the subject knowledge that is to be mastered.

While the above textbook analysis is admittedly impressionistic, it would be useful to find out when and how glosses ARE regularly used in real reading materials for L2 learners, because this could help to better align the research on this topic with real-world practices. I am not saying that research should necessarily imitate established practices, because trying out innovative procedures may reveal more effective alternatives. All the same, findings from research on glossing where it was implemented in ways that practitioners consider unrealistic (for example because the glossing looks overly excessive) are unlikely to make an impact beyond the research community (Boers, 2021, pp. 205–210; Marsden & Kasprowicz, 2017, for pleas to build stronger connections between the worlds of research and educational practice).

6. The future of glossing

Reading habits are changing rapidly, with onscreen reading becoming the default mode. With the advances in technology, materials developers no longer need to create glosses themselves, but can insert hyperlinks to online resources, such as online dictionaries (e.g., Godwin-Jones, 2018, for a discussion of the affordances of technology for L2 reading). Hulstijn et al. (1996) and Luppescu and Day (1993) found that students were reluctant to look up words in PRINT dictionaries. However, if a dictionary entry is just one mouse-click away, then more lookups might be expected. An interesting topic for research, then, would be to compare learning gains from hyperlinks to dictionaries and hyperlinks to tailor-made glosses. When inspecting a dictionary entry, the learner will usually need to evaluate which of the meanings listed there matches the contextual use of the item, which requires some cognitive effort, thus benefiting retention (e.g., Laufer & Hulstijn, 2001; but see Yanagisawa & Webb, 2021).

If providing hyperlinks is getting increasingly easy, then will such links become increasingly abundant in onscreen texts? If so, this raises questions about the impact of various glossing densities again. How often and under what circumstances will L2 readers make use of hyperlinks if these are

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

omnipresent? Will materials designers need to help learners select, by highlighting the items that they consider particularly relevant, assuming this will prompt lookups (e.g., De Ridder, 2002)?

A more general question that needs to be asked is to what extent the findings of past research on glosses in print materials will remain informative in an era where readers have numerous resources at their fingertips to help them understand L2 words and texts.

7. Beyond word meanings

Given the purpose of glossing is to assist reading comprehension, the information provided by glosses has typically been about the MEANING of words. Still, other aspects of word knowledge have occasionally been targeted. This includes pronunciation (Yeh & Wang, 2003), collocation (Lee et al., 2017), and grammar (Jung, 2016). It is not clear to what extent learners should be expected to engage with these aspects if their primary interest supposedly lies with the content of the texts instead of the language code proper. Jung (2016) tried to raise learners' appreciation of the syntactic characteristics of certain verbs by offering translations of their instances in the margin of the text, but this did not appear to have much effect, likely because the learners (understandably) thought these glosses were intended as assistance with text comprehension, not as a prompt for grammar learning. To make learners appreciate this other potential role of glosses, additional grammatical information could be signalled by means of meta-linguistic information (e.g., *drove* = past tense of *drive*).[15] Another option is to use typographic means to direct the learners' attention to specific elements within the gloss. This could be helpful, for example, in the case of glosses for collocations that differ between L1 and L2 (e.g., *have a nightmare* in English = *faire un cauchemar* in French). It is debatable whether the learning gains from such kinds of glossing should still be considered a by-product of content-focused reading, though, because it asks learners to focus on aspects of language that do not contribute much to text comprehension.

8. Interim conclusion

Although research on glossing has a long history and the available collection of research studies on the topic is large, many research questions have been answered in only tentative and provisional ways, in part owing to the wide diversity in research designs across studies—making meta-analytic endeavours challenging. When it comes to comparing the effectiveness of certain types of glossing (e.g., in-text vs. marginal; single-meaning vs. MC; single-mode vs. multimodal), the number of well-controlled and transparently reported studies is not yet large enough to draw more than preliminary conclusions. For now, the following conclusions do seem justified.

First, the difference in the amount of vocabulary learning from glossed and non-glossed text is likely to be very marked when the non-glossed version of the text poses serious comprehension problems, as the learners who are given this version will not stand a good chance of inferring meanings of unfamiliar words. However, giving a group of learners a text for unassisted reading that they are not ready for has questionable ecological validity.

Second, when zooming in on the benefits of glossed reading per se, the best learning outcomes as far as vocabulary gains are concerned tend to emerge when learners (a) access the glosses (instead of ignoring them) and do so more than once, (b) access the glosses and then re-encounter the words in the text, and (c) treat the glosses as study material rather than mere reading support. All this is under the assumption that (d) the quality of the glosses is such that they foster accurate understanding of the words' meanings, and (e) the learners find the amount of new information cognitively manageable.

9. Directions for future research

A lot of work clearly remains to be done, and an important part of this will be to follow more systematic and transparent research protocols to enhance the replicability and comparability of the studies

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

(Porte & McManus, 2019). For the sake of ecological validity, it will be important to consider how well the texts used in these studies match the proficiency of the learners. If the text matches the learners' proficiency level, then there cannot be an abundance of unfamiliar words, and so no excessive number of glosses will be required to make the text comprehensible. Researchers interested in the role of glossing density should nonetheless be able to design experiments with differing densities even if only 2–3% of the running words are unknown, provided the reading texts are long enough.

For the sake of establishing cause-effect relationships, it will be useful to systematically collect behavioural data as the participants read the texts and the glosses. This may be computer-log data regarding hyperlink use and/or eye-tracking data regarding attention given to different elements of the text and its glosses during reading. It will help to ascertain that the glosses were inspected by the participants (and how often and how long). In addition, data on how the participants subjectively experienced the reading task will be useful to help attribute learning gains to incidental acquisition rather than deliberate study. For example, even though they were not told a vocabulary quiz was going to follow, did they consider the glossed words as study material in their own right? Did learner-participants suspect the 'true' purpose of the reading activity was vocabulary acquisition?

Still underexplored to date as well is how individual learner characteristics, such as differences in learning style (e.g., Plass et al., 1998, 2003; Rassaei, 2018), age (e.g., Yeung, 1999), and working-memory capacity (e.g., Varol & Erçetin, 2019) moderate the effects of specific implementations of glossing.

Given the positive association between learning gains and repetition, scenarios where glossed words occur more than once in a text seem particularly promising. On which occurrence of the word in the text (first, second, third) is it best to provide a gloss? If a gloss is provided more than once for the same word, then on which occurrence of the word should it be reiterated? If a word is glossed more than once, will variation in the clarification (e.g., first pictorial, then textual; first L1 then L2; first literal meaning, then figurative meaning) be more beneficial than presenting the same clarification twice? Does a gloss that leaves some doubt about the precise meaning of a word arouse curiosity about subsequent instances of the word in the text? Does distributing target words over separate pages or screens after providing a gloss lead to better retention (owing to the effort to retrieve meanings from memory) than showing repeated words on the same page/screen as the gloss? These are all questions that have begun to attract attention in recent years, but that still invite more systematic inquiry to ensure that recommendations made for the design of L2 reading materials are well substantiated.

Questions arising

1. How large is the effect of glossing on vocabulary learning if it is compared with reading non-glossed texts at the right level of difficulty for the learner group, that is, texts that are comprehensible enough for meaning inferencing to be feasible?
2. How do text-based output activities influence learners' use of glosses and their retention of glossed items?
3. How large is the effect of glossing relative to reading non-glossed text if participants are explicitly informed the reading activity serves the purpose of vocabulary learning?
4. How does glossing density influence learners' use of glosses, their retention of glossed items, and their engagement with text content?
5. Are L1 glosses (i.e., translations) more beneficial for vocabulary learning than L2 glosses also if the latter are carefully designed to be concise and transparent for the learner group?
6. Do glosses that leave some ambiguity or vagueness regarding the contextual meaning of a lexical item prompt renewed interest in the item when it is re-encountered in the text, and does this lead to better retention?
7. How do the benefits of hyperlinked glosses compare with online dictionaries now the latter have become so easy to access?
8. What are the effects of adding information to glosses that goes beyond the clarification of meaning but instead draws attention to formal features (e.g., grammar, collocation) on learners' attention to (and retention of) the text content, the meaning of the glossed items, and the additional, form-related information?

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

Acknowledgements. I am grateful to the seven reviewers who gave feedback on a previous version of this review. Their insightful and detailed suggestions have greatly benefited its content and its readability.

Endnotes

[1] Abraham (2008), in fact, included seven primary studies that examined vocabulary learning from reading glossed texts, but I exclude one of them (Plass et al., 2003), because, unlike the others, it did not include an immediate posttest, which hinders a direct comparison. Even so, the remaining studies show wide diversity in design and so it is no surprise the outcomes vary considerably. For example, one of the studies (Knight, 1994) is, in fact, about lookups in an electronic dictionary rather than about the use of glosses. In one of the studies reporting a large effect of glossing (Bowles, 2004), the participants were asked to think aloud as they read the text. Many of the students who read the non-glossed version apparently misunderstood this as reading the text aloud, which did not help them to engage with either text content or word meanings. Their performance on text comprehension and vocabulary posttests was very poor, thus contributing to the wide gap between their scores and those of the participants who read the glossed version.

[2] The effect size for reading comprehension reported in Abraham (2008, p. 225) for one of the primary studies, Ben Salem (2006), is extraordinarily large: $g = 3.42$. According to my own effect size calculation based on the descriptive statistics (Ben Salem, 2006, p. 70), the first digit must be a typo, and should in fact be 2. This still leaves an exceptionally large effect. However, the text comprehension questions in this study were essentially about the meaning of the target words, and the text was deliberately made devoid of contextual clues to the words' meanings. The learners who were given the non-glossed version could therefore hardly be expected to pick up these meanings and do well on either the vocabulary test or the text comprehension test. Indeed, their performance on the multiple-choice questions in these tests was hardly above chance. That both reading comprehension and meaning inferencing were deliberately made extra hard for these participants helps to explain the wide gap between their test scores and the scores of the participants who read the glossed version of the text.

[3] This analysis concerns data sets from Hulstijn et al. (1996), Lee et al. (2016), Shiki (2008), Watanabe (1997) and Zhao & Ren (2019). I am very grateful to Akifumi Yanagisawa and his co-authors for making information about the primary studies included in their meta-analytic review available as an online supplement to their article. This level of transparency is exemplary.

[4] The case of vocabulary learning from reading glossed texts illustrates that INCIDENTAL learning (defined here as learning as a by-product of content-focused activities) is not synonymous to IMPLICIT learning, that is, learning without awareness on the part of the learners that they are learning something (Norbert Schmitt, personal communication).

[5] Half of the glossed words in Hulstijn et al. (1996) occurred three times in the text. The glossing density of 2.5% of the running words includes the repeated instances of these words, but the total number of marginal glosses for the 1,300-word text was just 16—a much lighter density than in many other studies, and similar to what I have seen in EFL textbooks.

[6] Kim et al. (2020) report $g = 0.80$ and $g = 0.15$ for 'beginner' and 'intermediate or higher' learners, respectively. These effect sizes reflect not only vocabulary learning but also text comprehension measures. How the proficiency levels were delineated is not explained.

[7] Many of the L2 definitions in studies such as Öztürk and Yorgancı (2017, pp. 647–650) seem unnecessarily elaborate. For example, 'data (n.): facts or information, especially when examined and used to find out things or to make decisions', 'gender (n.) the fact of being male or female, especially when considered with reference to social and cultural differences, not differences in biology', and alter (v.): to become different; to make somebody/something different.' Surely, these could have been more concise (e.g., 'alter' (v.): 'change').

[8] Kim et al. (2020; Table 5) found no compelling evidence of an effect of gloss density. However, several of the primary studies included in their meta-analysis lacked information about glossing density, leaving just a handful of data sets for this analysis. It also included studies that focused on text comprehension, not vocabulary learning.

[9] There is another meta-analysis claiming superiority of text-plus-picture glosses over single-mode glosses: Vahedi et al. (2016). An issue with this meta-analysis, however, is that, on examining the list of primary studies on which its conclusion is based, it transpires that several experiments were included that did not actually compare textual and pictorial glosses (e.g., De Ridder, 2002; Poole, 2012; Yoshii, 2013). This is puzzling because it contradicts the inclusion criteria stipulated by the authors. I will therefore refrain from saying more about Vahedi et al.'s meta-analysis in this article, apart from cautioning readers that meta-analytic reviews need to be interpreted cautiously, just like any other research report (e.g., Boers et al., 2020).

[10] Interestingly, Ramezanali et al. (2021) found that adding a THIRD gloss did not lead to significant additional vocabulary gains. However, it is not clear how many studies included in their meta-analysis had reading conditions that provided more than two glosses. Perhaps this subsample was simply not large enough for a trend to reach statistical significance.

[11] An interesting finding of Yoshii (2013) was that a considerable number of students who were given the text with access to single-meaning glosses (instead of multiple-choice glosses) made very little use of these. The author then decided to exclude these students' test data, arguing that it is the effect of a gloss WHEN IT IS ACCESSED by the learner that must be of interest. An alternative perspective is to consider the likelihood that a specific type of gloss attracts interest a feature that is inherent to the

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

glossing condition, with practical relevance for the design of reading materials. Had Yoshii (2013) included the test data from the students who ignored the glosses, then the conclusion that single-meaning glosses brought about better gains than the multiple-choice glosses might no longer have been substantiated.

[12] Rott (2007) also included a reading condition where the students were invited to recall the meaning of words that had been glossed on their first occurrence in the text. However, the students were not prevented from revisiting the glosses, and so it is not certain how often they really tried to retrieve word meanings from memory instead of simply looking at the gloss again.

[13] Jung and Révész (2018) did not include a measure of vocabulary learning. However, a follow-up study by Jung (2020), in which the same text-reordering tasks were used, did gauge learning, by means of a multiple-choice test. The average scores were not above chance, which is not very surprising if these learners—like in Jung and Révész (2018)—did not find the glosses useful to complete the task.

[14] The textbooks I examined were ones available in the library at my faculty. This convenience sample consisted of *Cutting Edge* (2013), *Interchange* (2017), *New American Inside Out* (2012), *New American Headway* (2016), *Straightforward* (2012), and *New Total English* (2011).

[15] Gettys et al. (2001) compared glosses that provide a translation of words specific to their inflectional use in the text (e.g., a verb in its past tense; e.g., DROVE) with the more standard glosses that provide a lemmatized translation (e.g., the verb in its infinitive form; e.g., DRIVE, regardless of whether the L2 word was inflected in the text), and found the latter to be more effective for learning the meanings of the words. However, Gettys et al. (2001) did not include tests to examine learning of the GRAMMATICAL properties of the words, and so it is impossible to tell whether the inflection-specific glosses had any effect in this regard.

References

Abraham, L. B. (2007). Second language reading comprehension and vocabulary learning with multimedia. *Hispania*, *90*(1), 98–109. https://doi.org/10.2307/20063468

Abraham, L. B. (2008). Computer-mediated glosses in second language reading comprehension and vocabulary learning: A meta-analysis. *Computer Assisted Language Learning*, *21*(3), 199–226. https://doi.org/10.1080/09588220802090246

Acha, J. (2009). The effectiveness of multimedia programmes in children's vocabulary learning. *British Journal of Educational Technology*, *40*(1), 23–31. https://doi.org/10.1111/j.1467-8535.2007.00800.x

Aizawa, K. (1998). Incidental vocabulary learning through reading: Guessing exercises, glossing or accessing dictionaries? *The Bulletin of the Kanto-Koshin-Etsu English Language Education Society*, *12*, 79–94. https://doi.org/10.20806/katejo.12.0_79

Akbulut, Y. (2007). Effects of multimedia annotations on incidental vocabulary learning and reading comprehension of advanced learners of English as a foreign language. *Instructional Science*, *35*(6), 499–517. https://doi.org/10.1007/s11251-007-9016-7

Al-Seghayer, K. (2001). The effect of multimedia annotation modes on L2 vocabulary acquisition: A comparative study. *Language Learning & Technology*, *5*(1), 202–232.

Arpaci, D. (2016). The effects of accessing L1 versus L2 definitional glosses on L2 learners' reading comprehension and vocabulary learning. *Eurasian Journal of Applied Linguistics*, *2*(1), 15–29. https://doi.org/10.32601/ejal.460988

Azari, F., Abdullah, F. S., Heng, C. S., & Hoon, T. B. (2012). *Effects of glosses on vocabulary gain and retention among tertiary level EFL learners*. Retrieved from https://eric.ed.gov/?id=ED533228

Barcroft, J. (2007). Effects of opportunities for word retrieval during second language vocabulary learning. *Language Learning*, *57*(1), 35–56. https://doi.org/10.1111/j.1467-9922.2007.00398.x

Barcroft, J. (2015). Can retrieval opportunities increase vocabulary learning during reading? *Foreign Language Annals*, *48*(2), 236–249. https://doi.org/10.1111/flan.12139

Ben Salem, E. (2006). *The influence of electronic glosses on word retention and reading comprehension with Spanish language learners. PhD dissertation*. Kansas University.

Bensoussan, M., & Laufer, B. (1984). Lexical guessing in context in EFL reading comprehension. *Journal of Research in Reading*, *7*(1), 15–32. https://doi.org/10.1111/j.1467-9817.1984.tb00252.x

Bernhardt, E. B. (2010). *Understanding advanced second language reading*. Routledge.

Boers, F. (2000). Enhancing metaphoric awareness in specialised reading. *English for Specific Purposes*, *19*(2), 137–147. https://doi.org/10.1016/S0889-4906(98)00017-9

Boers, F. (2021). *Evaluating second language vocabulary and grammar instruction: A synthesis of the research on teaching words, phrases, and patterns*. Routledge.

Boers, F., Bryfonski, L., Faez, F., & McKay, T. (2020). A call for cautious interpretation of meta-analytic reviews. *Studies in Second Language Acquisition*, *43*(1), 2–24. https://doi.org/10.1017/S0272263120000327

Boers, F., Dang, T. C. T., & Strong, B. (2017). Comparing the effectiveness of phrase-focused exercises. A partial replication of Boers, Demecheleer, Coxhead, and Webb (2014). *Language Teaching Research*, *21*(3), 362–280. https://doi.org/10.1177/1362168816651464

Boers, F., Warren, P., He, L., & Deconinck, J. (2017). Does adding pictures to glosses enhance vocabulary uptake from reading? *System*, *66*, 113–129. https://doi.org/10.1016/j.system.2017.03.017

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

Bowles, M. A. (2004). L2 glossing: To CALL or not to CALL. *Hispania*, *87*(3), 541–552. https://doi.org/10.2307/20063060
Cheng, Y.-H., & Good, R. L. (2009). L1 glosses: Effects on EFL learners' reading comprehension and vocabulary retention. *Reading in a Foreign Language*, *21*(2), 119–142.
Chun, D., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *Modern Language Journal*, *80*(2), 183–198. https://doi.org/10.1111/j.1540-4781.1996.tb01159.x
Davis, J. N. (1989). Facilitating effects of marginal glosses on foreign language reading. *Modern Language Journal*, *73*(1), 41–48. https://doi.org/10.1111/j.1540-4781.1989.tb05308.x
Day, R. R., Omura, C., & Hiramatsu, M. (1991). Incidental EFL vocabulary learning and reading. *Reading in a Foreign Language*, *7*(2), 541–551.
De Ridder, I. (2002). Visible or invisible links: Does the highlighting of hyperlinks affect incidental vocabulary learning, text comprehension, and the reading process? *Language Learning & Technology*, *6*(1), 123–146.
Elgort, I. (2017). Incorrect inferences and contextual word learning in English as a second language. *Journal of the European Second Language Association*, *1*(1), 1–11. https://doi.org/10.22599/jesla.3
Elgort, I., Beliaeva, N., & Boers, F. (2020). Contextual word learning in the first and second language: Definition placement and interference error effects on declarative and non-declarative knowledge. *Studies in Second Language Acquisition*, *42*(1), 7–32. https://doi.org/10.1017/S0272263119000561
Elgort, I., & Warren, P. (2014). L2 vocabulary learning from reading: Explicit and tacit lexical knowledge and the role of learner and item variables. *Language Learning*, *64*(2), 365–414. https://doi.org/10.1111/lang.12052
Ertürk, ZÖ. (2016). The effect of glossing on EFL learners' incidental vocabulary learning in reading. *Procedia: Social and Behavioral Sciences*, *232*, 373–381. https://doi.org/10.1016/j.sbspro.2016.10.052
Farvardin, M. T., & Biria, R. (2012). The impact of gloss types on Iranian EFL students' reading comprehension and lexical retention. *International Journal of Instruction*, *5*(1), 99–114.
Gan, X. (2014). Study on the effects of gloss type on Chinese EFL learners' incidental vocabulary acquisition. *Theory and Practice in Language Studies*, *4*(6), 1251–1256. https://doi.org/10.4304/tpls.4.6.1251-1256
Gettys, S., Imhof, L. A., & Kautz, J. O. (2001). Computer-assisted reading: The effect of glossing format on comprehension and vocabulary retention. *Foreign Language Annals*, *34*(2), 91–99. https://doi.org/10.1111/j.1944-9720.2001.tb02815.x
Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A., & Yoon, H.-J. (2018). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, *21*(3), 563–584. https://doi.org/10.1017/S1366728917000219
Godfroid, A., Boers, F., & Housen, A. (2013). An eye for words: Gauging the role of attention in L2 vocabulary acquisition by means of eye-tracking. *Studies in Second Language Acquisition*, *35*(3), 483–517. https://doi.org/10.1017/S0272263113000119
Godfroid, A., & Schmidtke, J. (2013). What do eye movements tell us about awareness? A triangulation of eye-movement data, verbal reports and vocabulary learning scores. In J. M. Bergsleithner, S. N. Frota, & J. K. Yoshioka (Eds.), *Noticing and second language acquisition: Studies in honour of Richard Schmidt* (pp. 183–205). University of Hawai'i.
Godfroid, A., & Spino, L. A. (2015). Reconceptualizing reactivity of think-alouds and eye tracking: Absence of evidence is not evidence of absence. *Language Learning*, *65*(4), 896–928. https://doi.org/10.1111/lang.12136
Godwin-Jones, R. (2018). Contextualized vocabulary learning. *Language Learning & Technology*, *22*(3), 1–19.
Holley, F. M., & King, J. K. (1971). Vocabulary glosses in foreign language reading materials. *Language Learning*, *21*(2), 213–219. https://doi.org/10.1111/j.1467-1770.1971.tb00060.x
Horst, M., Cobb, T., & Meara, P. (1998). Beyond A Clockwork Orange: Acquiring second language vocabulary through reading. *Reading in a Foreign Language*, *11*(2), 207–223.
Hu, H.-C. M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*(1), 403–430.
Huang, L.-L., & Lin, C.-C. (2014). Three approaches to glossing and their effects on vocabulary learning. *System*, *44*, 127–136. https://doi.org/10.1016/j.system.2014.03.006
Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental vocabulary learning. In P. J. L. Arnaud, & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (pp. 113–125). Palgrave Macmillan.
Hulstijn, J. H. (2003). Incidental and intentional learning. In C. Doughty, & M. Long (Eds.), *The handbook of second language acquisition* (pp. 349–381). Blackwell Publishing.
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*(3), 327–339. https://doi.org/10.2307/329439
Jacobs, G. M. (1994). What lurks in the margin: Use of vocabulary glosses as a strategy in second language reading. *Issues in Applied Linguistics*, *5*(1), 115–137. https://doi.org/10.5070/L451005174
Jacobs, G. M., Dufon, P., & Hong, F. C. (1994). L1 and L2 vocabulary glosses in L2 reading passages: Their effectiveness for increasing comprehension and vocabulary knowledge. *Journal of Research in Reading*, *17*(1), 19–28. https://doi.org/10.1111/j.1467-9817.1994.tb00049.x
Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *Modern Language Journal*, *86*(4), 546–561. https://doi.org/10.1111/1540-4781.00160

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

Jung, J. (2016). Effects of glosses on learning of L2 grammar and vocabulary. *Language Teaching Research*, *20*(1), 92–112. https://doi.org/10.1177/1362168815571151

Jung, J. (2020). The impact of glossing and reading activity manipulation on learning of L2 lexico-grammatical and lexical items. *Language Teaching Research*, Online First. https://doi.org/10.1177/1362168820911198

Jung, J., & Révész, A. (2018). The effects of reading activity characteristics on L2 reading processes and noticing of glossed constructions. *Studies in Second Language Acquisition*, *40*(4), 755–780. https://doi.org/10.1017/S0272263118000165

Kang, H., Kweon, S.-O., & Choi, S. (2020). Using eye-tracking to examine the role of first and second language glosses. *Language Teaching Research*, Online First. https://doi.org/10.1177/1362168820928567

Khezrlou, S., Ellis, R., & Sadeghi, K. (2017). Effects of computer-assisted glosses on EFL learners' vocabulary acquisition and reading comprehension in three learning conditions. *System*, *65*, 104–116. https://doi.org/10.1016/j.system.2017.01.009

Kim, H. S., Lee, J. H., & Lee, H. (2020). The relative effects of L1 and L2 glosses on L2 learning: A meta-analysis. *Language Teaching Research*, Online First. https://doi.org/10.1177/1362168820981394

Knight, S. (1994). Dictionary use while reading: The effects on comprehension and vocabulary acquisition for learners of different verbal abilities. *Modern Language Journal*, *78*(3), 285–299. https://doi.org/10.1111/j.1540-4781.1994.tb02043.x

Ko, H. M. (1995). Glossing in incidental and intentional learning of foreign language vocabulary and reading. *University of Hawai'i Working Papers in ESL*, *13*(2), 49–94.

Ko, M. H. (2012). Glossing and second language vocabulary learning. *TESOL Quarterly*, *46*(1), 56–79. https://doi.org/10.1002/tesq.3

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge University Press.

Kost, C. R., Foss, P., & Lenzini, J. J. (1999). Textual and pictorial glosses: Effectiveness on incidental vocabulary growth when reading in a foreign language. *Foreign Language Annals*, *32*(1), 89–97. https://doi.org/10.1111/j.1944-9720.1999.tb02378.x

Laufer, B. (2020). Lexical coverages, inferencing unknown words and reading comprehension: How are they related? *TESOL Quarterly*, *54*(4), 1076–1085. https://doi.org/10.1002/tesq.3004

Laufer, B., & Hulstijn, J. H. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics*, *22*(1), 1–26. https://doi.org/10.1093/applin/22.1.1

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*(1), 15–30.

Laufer, B., & Sim, D. D. (1985). Taking the easy way out: Non-use and misuse of contextual clues in EFL reading comprehension. *English Teaching Forum*, *18*(2), 405–411.

Lee, H., Lee, H., & Lee, J. H. (2016). Evaluation of electronic and paper textual glosses on second language vocabulary learning and reading comprehension. *The Asia-Pacific Education Researcher*, *25*(4), 499–507. https://doi.org/10.1007/s40299-015-0270-1

Lee, H., Warschauer, M., & Lee, J. H. (2017). The effects of concordance-based electronic glosses on L2 vocabulary learning. *Language Learning & Technology*, *21*(2), 32–51.

Lomicka, L. (1998). To gloss or not to gloss: An investigation of reading comprehension online. *Language Learning & Technology*, *1*(2), 41–50.

Luppescu, S., & Day, R. R. (1993). Reading, dictionaries, and vocabulary learning. *Language Learning*, *43*(2), 263–279. https://doi.org/10.1111/j.1467-1770.1992.tb00717.x

Marsden, E., & Kasprowicz, R. (2017). Foreign language educators' exposure to research: Reported experiences, exposure via citations, and a proposal for action. *Modern Language Journal*, *101*(4), 613–642. https://doi.org/10.1111/modl.12426

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). Cambridge University Press.

Mayer, R. E., & Anderson, R. W. (1992). The instructive animation: Helping students build connections between words and pictures in multimedia learning. *Journal of Educational Psychology*, *84*(4), 444–452. https://doi.org/10.1037/0022-0663.84.4.444

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, *82*(4), 715–726. https://doi.org/10.1037/0022-0663.82.4.715

Mitarai, J., & Aizawa, K. (1999). The effects of different types of glosses in vocabulary learning and reading comprehension. *ARELE: Annual Review of English Language Education in Japan*, *10*, 73–82. https://doi.org/10.20581/arele.10.0_73

Miyasako, N. (2002). Does text-glossing have any effects on incidental vocabulary learning through reading for Japanese senior high school students? *Language Education & Technology*, *39*, 1–20. https://doi.org/10.24539/let.39.0_1

Mondria, J. (2003). The effects of inferring, verifying, and memorizing on the retention of L2 word meanings: An experimental comparison of the "meaning-inferred method" and the "meaning-given method.". *Studies in Second Language Acquisition*, *25*(4), 473–499. https://doi.org/10.1017/S0272263103000202

Montero Perez, M., Peters, E., & Desmet, P. (2018). Vocabulary learning through viewing video: The effect of two enhancement techniques. *Computer Assisted Language Learning*, *31*(1–2), 1–26. https://doi.org/10.1080/09588221.2017.1375960

Moradan, A., & Vafaei, M. (2016). The effect of glosses on incidental vocabulary learning of Iranian EFL learners. *International Journal of Applied Linguistics & English Literature*, *5*(6), 34–42. https://doi.org/10.7575/aiac.ijalel.v.5n.6p.34

Na, L., & Nation, I. S. P. (1985). Factors affecting guessing vocabulary in context. *RELC Journal*, *16*(1), 33–42. https://doi.org/10.1177/003368828501600103

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

Nagata, N. (1999). The effectiveness of computer-assisted interactive glosses. *Foreign Language Annals*, *32*(4), 469–479. https://doi.org/10.1111/j.1944-9720.1999.tb00876.x

Nassaji, H. (2003). L2 vocabulary learning from context: Strategies, knowledge sources, and their relationship with success in L2 lexical inferencing. *TESOL Quarterly*, *37*(4), 645–670. https://doi.org/10.2307/3588216

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Ouyang, J., Huang, L., & Jiang, J. (2020). The effects of glossing on incidental vocabulary learning during second language reading: Based on an eye-tracking study. *Journal of Research in Reading*, *43*(4), 496–515. https://doi.org/10.1111/1467-9817.12326

Öztürk, M. S., & Yorgancı, M. (2017). Effects of L1 and L2 glosses on incidental vocabulary learning of EFL prep students. *International Periodical for the Languages, Literature and History of Turkish or Turkic*, *12*(6), 635–656. http://dx.doi.org/10.7827/TurkishStudies.11432

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford University Press.

Pellicer-Sánchez, A. (2016). Incidental L2 vocabulary acquisition *from* and *while* reading. *Studies in Second Language Acquisition*, *38*(1), 97–130. https://doi.org/10.1017/S0272263115000224

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*(1), 113–151. https://doi.org/10.1111/j.1467-9922.2009.00502.x

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*(2), 271–275.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology*, *90*(1), 25–36. https://doi.org/10.1037/0022-0663.90.1.25

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior*, *19*(2), 221–243. https://doi.org/10.1016/S0747-5632(02)00015-8

Porte, G., & McManus, K. (2019). *Doing replication research in applied linguistics*. Routledge.

Poole, R. (2012). Concordance-based glosses for academic vocabulary acquisition. *CALICO Journal*, *29*(4), 679–693. https://doi.org/10.11139/cj.29.4.679-693

Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language Learning*, *54*(3), 469–523. https://doi.org/10.1111/j.0023-8333.2004.00263.x

Qian, D. D., & Lin, L. H. F. (2020). The relationship between vocabulary knowledge and language proficiency. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 66–80). Routledge.

Ramezanali, N., Uchihara, T., & Faez, F. (2021). Efficacy of multimodal glossing on second language vocabulary learning: A meta-analysis. *TESOL Quarterly*, *55*(1), 105–133. https://doi.org/10.1002/tesq.579

Rassaei, E. (2018). Computer-mediated textual and audio glosses, perceptual style and L2 vocabulary learning. *Language Teaching Research*, *22*(6), 657–675. https://doi.org/10.1177/1362168817690183

Rott, S. (2005). Processing glosses: A qualitative exploration of how form-meaning connections are established and strengthened. *Reading in a Foreign Language*, *17*(2), 95–124.

Rott, S. (2007). The effect of frequency of input-enhancements on word learning and text comprehension. *Language Learning*, *57*(2), 165–199. https://doi.org/10.1111/j.1467-9922.2007.00406.x

Rott, S., Williams, J., & Cameron, R. (2002). The effect of multiple-choice L1 glosses and input-output cycles on lexical acquisition and retention. *Language Teaching Research*, *3*(3), 183–222. https://doi.org/10.1191/1362168802lr108oa

Rouhi, A., & Mohebbi, H. (2012). The effect of computer assisted L1 and L2 glosses on L2 vocabulary learning. *The Journal of Asia TEFL*, *9*(2), 1–19.

Sadeghi, K., & Ahmadi, N. (2012). The effect of gloss type and mode on Iranian EFL learners' vocabulary acquisition. *Issues in Language Teaching*, *1*(1), 159–188.

Salehi, V., & Naserieh, F. (2013). The effects of verbal glosses on vocabulary learning and reading comprehension. *Asian EFL Journal*, *15*(2), 24–64.

Salimi, E. A., & Mirian, E. S. (2019). The effect of L1 glosses for abstract words on English reading comprehension. *The Reading Matrix: An International Online Journal*, *19*(1), 181–196.

Samian, H. V., Foo, T. C. V., & Mohebbi, H. (2016). The effects of giving and receiving marginal L1 glosses on L2 vocabulary learning by upper secondary learners. *English Language Teaching*, *9*(2), 66–76. https://doi.org/10.5539/elt.v9n2p66

Schmidt, R. W. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3–32). Cambridge University Press.

Schmitt, N., Grabe, W., & Jiang, X. (2011). The percentage of words known in a text and reading comprehension. *Modern Language Journal*, *95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55–88. https://doi.org/10.1177/026553220101800103

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

Shahrokni, S. A. (2009). Second language incidental vocabulary learning: The effect of online textual, pictorial, and textual pictorial glosses. *The Electronic Journal for English as a Second Language*, *13*(3), 1–27.
Sharwood-Smith, M. (1993). Input enhancement in instructed SLA: Theoretical bases. *Studies in Second Language Acquisition*, *15*(2), 165–179. https://doi.org/10.1017/S0272263100011943
Shiki, O. (2008). Effects of glosses on incidental vocabulary learning: Which gloss-type works better, L1, L2, single choice, or multiple choices for Japanese university students? *Journal of Inquiry and Research*, *87*, 39–56.
Strong, B., & Boers, F. (2019). The error in trial and error: Exercises on phrasal verbs. *TESOL Quarterly*, *53*(2), 289–319. https://doi.org/10.1002/tesq.478
Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5
Taheri, P., & Zade, M. H. (2014). On the incidental learning of abstract and concrete vocabulary items in L1 gloss and L2 gloss. *English Language Teaching*, *1*(4), 77–91.
Teng, F. (2020). Retention of new words learned incidentally from reading: Word exposure frequency, L1 marginal glosses, and their combination. *Language Teaching Research*, *24*(6), 785–812. https://doi.org/10.1177/1362168819829026
Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, *69*(3), 559–599. https://doi.org/10.1111/lang.12343
Vahedi, V. S., Ghonsooly, B., & Pishghadam, R. (2016). Vocabulary glossing: A meta-analysis of the relative effectiveness of different gloss types on L2 vocabulary acquisition. *Teaching English With Technology*, *16*(1), 3–25.
Van den Broek, G. S. E., Takashima, A., Segers, E., & Verhoeven, L. (2018). Contextual richness and word learning: Context enhances comprehension but retrieval enhances retention. *Language Learning*, *68*(2), 546–585. https://doi.org/10.1111/lang.12285
Varol, B., & Erçetin, G. (2016). Effects of working memory and gloss type on L2 text comprehension and vocabulary learning in computer-based learning. *Procedia: Social and Behavioral Sciences*, *232*, 759–768. https://doi.org/10.1016/j.sbspro.2016.10.103
Varol, B., & Erçetin, G. (2019). Effects of gloss type, gloss position, and working memory capacity on second language comprehension in electronic reading. *Computer Assisted Language Learning*, Online First. https://doi.org/10.1080/09588221.2019.1643738
Warren, P., Boers, F., Grimshaw, G., & Siyanova-Chanturia, A. (2018). The effect of gloss type on learners' intake of new words during reading: Evidence from eye-tracking. *Studies in Second Language Acquisition*, *40*(4), 883–906. https://doi.org/10.1017/S0272263118000177
Watanabe, Y. (1997). Input, intake, and retention: Effects of increased processing on incidental learning of foreign language vocabulary. *Studies in Second Language Acquisition*, *19*(3), 287–307. https://doi.org/10.1017/S027226319700301X
Webb, S. (2020). Incidental vocabulary acquisition. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge.
Webb, S., Sasao, Y., & Balance, O. (2017). The updated vocabulary levels tests: Developing and validating two forms of the VLT. *ITL – International Journal of Applied Linguistics*, *168*(1), 33–69. https://doi.org/10.1075/itl.168.1.02web
Yanagisawa, A., & Webb, S. (2021). To what extent does the Involvement Load Hypothesis predict incidental L2 vocabulary learning? A meta-analysis. *Language Learning*, *71*(2), 487–536. https://doi.org/10.1111/lang.12444
Yanagisawa, A., Webb, S., & Uchihara, T. (2020). How do different forms of glossing contribute to L2 vocabulary learning from reading? A meta-regression analysis. *Studies in Second Language Acquisition*, *42*(2), 411–438. https://doi.org/10.1017/S0272263119000688
Yanguas, I. (2009). Multimedia glosses and their effect on L2 text comprehension and vocabulary learning. *Language Learning & Technology*, *13*(2), 48–67.
Yeh, Y., & Wang, C. (2003). Effects of multimedia vocabulary annotations and learning styles acquisition. *CALICO Journal*, *21*(1), 131–144. https://doi.org/10.1558/cj.v21i1.131-144
Yeung, A. S. (1999). Cognitive load and learner expertise: Split-attention and redundancy effects in reading comprehension tasks with vocabulary definitions. *The Journal of Experimental Education*, *67*(3), 197–217. https://doi.org/10.1080/00220979909598353
Yoshii, M. (2006). L1 and L2 glosses: Their effects on incidental vocabulary learning. *Language Learning & Technology*, *10*(3), 85–101.
Yoshii, M. (2013). Effects of gloss types on vocabulary learning through reading: Comparison of single and multiple gloss gloss types. Special issue on learner-computer interaction in language education: A Festschrift in honor of Robert Fisher, *CALICO Journal*, 203–229.
Yoshii, M., & Flaitz, J. (2002). Second language incidental vocabulary retention: The effect of text and picture annotation types. *CALICO Journal*, *20*(1), 33–58. https://doi.org/10.1558/cj.v20i1.33-58
Yun, J. (2011). The effects of hypertext glosses on L2 vocabulary acquisition: A meta-analysis. *Computer Assisted Language Learning*, *24*(1), 39–58. https://doi.org/10.1080/09588221.2010.523285
Zahar, R., Cobb, T., & Spada, N. (2001). Acquiring vocabulary through reading: Effects of frequency and contextual richness. *The Canadian Modern Language Review*, *57*(4), 541–572. https://doi.org/10.3138/cmlr.57.4.541

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press

Zarei, A. A., & Hasani, S. (2011). The effects of glossing conventions on L2 vocabulary recognition and production. *The Journal of Teaching Language Skills (JTLS)*, *3*(2), 209–233.
Zhang, Z., & Webb, S. (2019). The effects of reading bilingual books on vocabulary learning. *Reading in a Foreign Language*, *31*(1), 108–139.
Zhao, T., & Ren, J. (2019). Incidental L2 lexical acquisition in reading: The role of L2-gloss frequency and learner proficiency. *The Language Learning Journal*, *47*(5), 608–624. https://doi.org/10.1080/09571736.2017.1349168

Frank Boers is a Professor in Applied Linguistics and TESOL at the University of Western Ontario, Canada. His initial research interests (in the 1990s) concerned lexicology, especially metaphorical expressions. His current research interests, however, were sparked by his extensive experience as an EFL teacher and teacher trainer, and since the 2000s he has published mostly about instructed second/foreign language acquisition. His latest book is *Evaluating second language vocabulary and grammar instruction: A synthesis of the research on teaching words, phrases, and patterns* (Routledge, 2021).

Cite this article: Boers, F. (2022). Glossing and vocabulary learning. *Language Teaching*, *55*(1), 1–23. https://doi.org/10.1017/S0261444821000252

https://doi.org/10.1017/S0261444821000252 Published online by Cambridge University Press