Intentional and Incidental Vocabulary Learning: The Role of Historical Linguistics in the Second Language Classroom

JAMES M. STRATTON
The University of British Columbia, Department of English Language and Literatures, 397–1873 East Mall, Vancouver, BC V6T 1Z1, Canada E-mail: james.stratton@ubc.ca

Although English and German are both Germanic languages, due to various historical changes, many of their cognates are no longer easily recognizable. This study examined whether knowledge of language history can be beneficial to learners when learning English–German cognates. Thirty-five English-speaking second language (L2) learners of 3rd-semester German at an American university were assigned to either an intentional or incidental learning condition. The intentional group received explicit instruction on 2 historical sounds shifts (Second Germanic Sound Shift, Ingvaeonic Palatalization) and relevant historical semantic changes. In contrast, the incidental group carried out various activities that exposed learners to cognates through reading and oral communication tasks. Results indicate that the intentional group significantly outperformed the incidental group, suggesting that knowledge of language history can be beneficial to learners when learning historically related languages. Declarative knowledge of the historical changes led to significantly greater vocabulary gains and it also provided learners with a tool kit to correctly predict the meaning of several cognates they had not previously encountered before. This study has broad implications for vocabulary learning, language teaching, and applied historical linguistics.

Keywords: vocabulary learning; explicit instruction; instructed second language acquisition; applied historical linguistics; German

A LONG-STANDING QUESTION IN instructed second language (L2) acquisition is whether L2s are best learned implicitly or explicitly. While the majority of research suggests that L2 grammar rules are more amenable to explicit, as opposed to implicit learning conditions (Goo et al., 2015; Kang et al., 2019; Norris & Ortega, 2000; Spada & Tomita, 2010), it is not entirely clear how generalizable these results are to other linguistic domains, such as L2 vocabulary. Although the majority of first language (L1) vocabulary is acquired incidentally (Nation, 2013; Webb & Nation, 2017), incidental L2 vocabulary acquisition is generally less successful (Carpenter et al., 2012; Laufer, 2005). While L2 vocabulary can be acquired incidentally through, for instance, free voluntary reading (Krashen, 2004, 2011) or extensive reading (Nation, 2015; Webb & Chang, 2015a, 2015b), it is estimated that learners need to know approximately 95–98% of words in a text to successfully be able to infer the meaning of unknown words (Hu & Nation, 2000; Laufer, 1997). Multiple exposures and sufficiently rich contexts are also usually necessary for successful incidental vocabulary acquisition to take place (Hu, 2013; Rott, 1999; Webb, 2007). Although other methods, such as gaming (Ranalli, 2008; Sundqvist, 2019) and

The Modern Language Journal, 106, 4, (2022)
DOI: 10.1111/modl.12805
0026-7902/22/837–857 $1.50/0

television (Feng & Webb, 2020; Peters & Webb, 2018; Rodgers & Webb, 2020), can promote incidental acquisition, for practical reasons these activities usually take place outside of the language classroom (De Wilde et al., 2019; Thornbury, 2016).

Intentional learning conditions, however, can offer a number of advantages when combined with explicit instruction (Elgort, 2011; Laufer, 2005; Schmitt, 2008; Yamamoto, 2014). Because English and German are both Germanic languages, they share a large number of cognates, where cognates are defined as words that can be traced back to the same ancestral form (e.g., English "hound" and German "Hund" both come from the Germanic "*hundaz"[1]). However, due to various historical changes, such as the Second Germanic Sound Shift (Salmons, 2012, pp. 112–118) and Ingvaeonic Palatalization (Lass, 1994, p. 55), many English–German cognates are no longer easily recognizable to naïve speakers of English. One of the changes triggered by the Second Germanic Sound Shift was the shift from the Germanic interdental fricative (Modern English "th") to the stop consonant [d] at the beginning of a word in German.[2] This sound change accounts for differences in English–German cognate pairings such as "Ding"–"thing," "Dorn"–"thorn," and "Durst"–"thirst."

Historical changes in meaning also account for many cognate differences between English and German. For instance, the German verb "sterben [to die]" is cognate with the English verb "to starve." Although "starve" originally meant "to die" (Old English "steorfan [to die]"), throughout the history of the English language, its meaning narrowed semantically to refer to a specific type of death (i.e., death by hunger). While associative and semantic elaboration (Craik & Lockhart, 1972; Craik & Tulving, 1975) have been shown to have positive effects on memory and learning (Bolger & Zapata, 2011; Hulstijn, 2001; Kirsch, 2012; McNamara & Scott, 2001; Prince, 2012) and scholars have been calling for the use of historical instruction in the German L2 classroom for quite some time (Horsford, 1987; Lightfoot, 2007; Smith, 1968; Wolff, 1993), to date, no empirical studies have tested the effects of explicit historical instruction on the learning of English–German cognates. Declarative knowledge of these historical changes may allow English-speaking L2 learners of German to create an instant connection between English and German cognates, while also providing them with a tool kit to correctly predict the meaning of cognates they have not encountered before. To address these gaps, the present study explores the effects of receiving historical instruction on cognate identification and predictability in the English-speaking L2 German classroom. Two research questions were posited:

RQ1. Based on an isolated translation task, is there a statistically significant difference between the number of cognates acquired by L2 learners who received historical instruction (intentional condition) and L2 learners who did not (incidental condition)?

H1. Given the positive effects of elaboration and association building on memory and learning (Laufer & Hulstijn, 2001; McNamara & Scott, 2001; Prince, 2012), the intentional condition is hypothesized to outperform the incidental condition because it will have encoded additional features to the memory trace (i.e., the historical information).

RQ2. Based on an isolated translation task, is there a statistically significant difference between the two learning conditions (intentional and incidental) in the number of German cognates L2 learners were able to correctly predict the meaning of? (Unlike with RQ1, these are cognates that L2 learners will not encounter in their pedagogical intervention.)

H2. The intentional condition is hypothesized to outperform the incidental condition because the intentional condition will have a tool kit (i.e., declarative knowledge of historical sound changes) through which the English cognate counterpart can be inferred.

BACKGROUND

Terminology

Intentional and incidental learning can be defined in different ways in the literature (Webb, 2020). One way to define intentional learning is by the intention of the learner. Learning words from a vocabulary list for a test would be an example of intentional vocabulary learning since the words are learned as a direct consequence of the learner's intention. In contrast, incidental learning refers to learning that takes place without the intention to do so. For instance, if learners read for pleasure and acquire the meaning of words as a result, this process would be an example of incidental vocabulary learning since their initial

intention was not to learn vocabulary. In the words of Gass (1999), incidental vocabulary learning is "a by-product of other cognitive exercises (e.g. reading/listening) involving comprehension" (p. 318).

Related to this distinction is the difference between explicit and implicit learning: Explicit learning can be defined as learning that takes place "with metalinguistic awareness" and implicit learning can be defined as learning that takes place "without metalinguistic awareness" (Ellis, 2009, p. 7).[3] Similarly, explicit instruction refers to instruction where the goal of the lesson (e.g., to learn particular words) is made overtly clear to the learner, whereas, in implicit instruction, the goal of the lesson is not made overtly clear to the learner (Ellis, 2008). Explicit instruction also often involves the use of metalanguage, whereas implicit instruction typically does not (Roehr, 2018, p. 64). Although the two learning conditions in this study are described as intentional versus incidental, a major difference between this study and previous ones is that the intentional group also received explicit instruction on relevant historical changes. In other words, learners in the intentional condition had the intention to learn the meaning of the cognates, but they were also given explicit strategies on how to do so.[4]

Incidental and Intentional Vocabulary Learning

Since most L1 vocabulary is acquired incidentally (Nation, 2013; Webb & Nation, 2017), the large majority of L2 vocabulary research has focused on incidental vocabulary acquisition through reading (Carpenter et al., 2012; Hulstijn, 1992; Laufer & Hulstijn, 2001; Rott, 1999; Webb & Nation, 2017). While L2 vocabulary can be acquired incidentally, composite vocabulary gains are generally reported to be minimal (Carpenter et al., 2012; Laufer, 2005), and incidental vocabulary acquisition is thought to be an incremental and gradual process (Webb & Nation, 2017). In contrast, intentional vocabulary learning is thought to be much quicker because explicit learning strategies can be used to ease the "learning burden" (Nation, 2013, p. 13). Through strategies such as elaboration (Barcroft, 2002; Craik & Lockhart, 1972; Craik & Tulving, 1975; Craik & Watkins, 1973), additional features can be encoded to a memory trace to make vocabulary more memorable and easier to retrieve (Gerrig & Zimbardo, 2010). In cases where the learner's L2 is related to the learner's L1, explicit connections between L1 and L2 vocabulary items may accelerate the acquisition process because novel stimuli are attached to representations that already exist (Hall, 2002).

Human Memory and Learning

For learning to take place, it is often thought that there must be a transfer of information from short-term to long-term memory (Atkinson & Shiffrin, 1968). Rehearsal is an encoding mechanism through which this serial transfer is thought to take place (Atkinson & Shiffrin, 1968; Craik & Tulving, 1975; Craik & Watkins, 1973). According to Craik & Watkins (1973), there are two types of rehearsal: maintenance rehearsal and elaborative rehearsal. Maintenance rehearsal is the conscious attempt to maintain information in short-term memory through, for instance, the phonological loop, whereas elaborative rehearsal is a metacognitive strategy that encodes additional features to a memory trace in an attempt to make it more memorable and distinctive (Gerrig & Zimbardo, 2010). Unlike maintenance rehearsal, only elaborative rehearsal (henceforth, elaboration) is thought to result in a transfer to long-term memory (Baddeley, 1990; Craik, 2002; Craik & Tulving, 1975) because elaboration involves deep processing whereas maintenance rehearsal involves shallow processing (Craik & Lockhart, 1972). In the context of vocabulary learning, this hypothesis would suggest that the more a learner pays attention to a word, the higher the likelihood that it will be retained (Laufer & Hulstijn, 2001).

A common way to elaborate is through association building, that is, the creation of a link between a novel stimulus and information already stored in long-term memory (Bolger & Zapata, 2011; Bower & Clark, 1969; Craik & Lockhart, 1972; Kirsch, 2012; McNamara & Scott, 2001; Prince, 2012). With respect to vocabulary learning, this can mean building an association between L1 and L2 vocabulary items. One way to create this link is through the use of narratives. For instance, in their seminal study, Bower and Clark (1969) asked 24 participants to learn 12 lists of 10 nouns. The participants were divided into two groups: a narrative group and a nonnarrative group. The nonnarrative group was asked to memorize the list without any instructions, and the narrative group was asked to memorize the words by putting them into a meaningful narrative. The nonnarrative group recalled only 13% of the words, whereas the narrative group recalled 93% of the words.

Although it is not yet clear on the basis of previous literature (or lack thereof) whether historical narratives, such as those provided in

the present study, yield the same results, fictional narrative and associative learning strategies have been the foundation for widely used memory techniques for centuries, such as the method of loci (Yates, 1966), mnemonics (Worthen & Hunt, 2011), and vocabulary learning strategies such as the keyword method (Atkinson & Raugh, 1975).[5] Drawing upon previous work on elaboration, narrative learning, and associative learning, one can hypothesize that receiving semantic explanations on the history of English–German cognates will have a significant effect on their learning.

Historical Instruction

In 1968, Smith published an article entitled "Historical Linguistics and the Teaching of German," in which he outlined various historical changes that may aid L2 German vocabulary learning. Among these changes were the Second Germanic Sound Shift and various semantic shifts. Since then, several scholars have echoed these claims (Crawford, 1988; Horsford, 1987; Lightfoot, 2007; Wolff, 1993), suggesting that historical instruction would placate the feeling of "foreignness." While Arteaga & Herschensohn (1995) found that explicit historical instruction on the French circumflex aided the learning of English–French loanwords, to date, no empirical studies have examined the effects of receiving historical instruction on the learning of English–German cognates.

Relationship Between English and German

Historical Changes. While the relationship between some English and German cognates is clear (e.g., "arm"–"Arm," "hand"–"Hand"), the relationship for others is less recognizable due to various historical sound (e.g., "strecken [to stretch]") and semantic changes (e.g., "Zaun [fence]"–"town"). Although English-speaking L2 learners of German might not immediately associate the German word "Zimmer [room]" with the English cognate counterpart "timber," both are descendants of the Germanic "*timbra [timber]" (Kroonen, 2013, p. 517). Throughout the history of the German language, the meaning of *timbra changed to refer to a specific place within a building (i.e., "Zimmer [room]"). However, the former meaning of 'timber,' which is still retained in English, lives on in German words such as "Zimmermann [carpenter]," an individual who works with wood. In terms of sound change, the difference between "Zimmer" and "timber" can be explained through a series of predictable, systematic, phonological developments that took place in German but not in English. The /t/ in "timber" systematically became /t͡s/ in word-initial position in German (represented by "z" orthographically), a phonological change that is part of a series of sound changes known as the Second Germanic Sound Shift (Salmons, 2012, pp. 112–118).[6] As for the change from "mb" to "mm," this is a type of assimilation that took place between Middle High German (ca. 1050–1350 CE) and Early New High German (ca. 1350–1650 CE).[7]

Second Germanic Sound Shift. Unlike the First Germanic Sound Shift, which differentiated Germanic languages from other Indo-European languages (Salmons, 2012, pp. 37–50), the Second Germanic Sound Shift differentiated Low German (e.g., English, Dutch, and Frisian) from Upper German dialects (e.g., Modern Standard German; cf. Salmons, 2012). Upper German voiceless stops /p, t, k/ became / p͡f, t͡s, kx̄/ at the beginning of a word and before a consonant (Salmons, 2012). The change from [p] to [p͡f] explains differences in cognates such as "pound"–"Pfund," "pipe"–"Pfeife," "pepper"–"Pfeffer," and "pan"–"Pfanne." The change from [t] to [t͡s] explains differences such as "tongue"–"Zunge," "too"–"zu," "twelve"–"zwölf," and "twig"–"Zweige," where, as aforementioned, the /t͡s/ sound is represented by "z" orthographically in German. As part of the Second Germanic Sound Shift, these German affricates became fricatives between vowels, accounting for differences such as "hope"–"hoffen," "help"–"helfen," and "water"–"Wasser"; and after vowels in final position, as in "hate"–"Hass" and "nut"–"Nuss." The change from the Germanic interdental fricative /θ-ð/ (Modern English "th") to German /d/ is also considered a possible further chain reaction of the Second Germanic Sound Shift (Stedje, 2001, p. 61). This change accounts for word differences such as "thing"–"Ding," "that"–"das," "thick"–"dick," "thorn"–"Dorn," and "thirst"–"Durst."

Ingvaeonic Palatalization. In contrast to the Second Germanic Sound Shift, which did not affect English, Ingvaeonic Palatalization affected English but not German (Barbour & Stevenson, 1990, pp. 31–32). This sound change describes the change from [k] to [ʧ]—that is, from "k" as in "cat," and "ch" as in "chair"—before front vowels.[8] This sound change accounts for differences in cognates such as "chin"–"Kinn," "child"–"Kind," "church"–"Kirche," "cheese"–"Käse," "crutch"–"Krücke," and "stretch"–"strecken."

Semantic Changes. In general, the meaning of a word can narrow or broaden over time (narrowing or broadening), it can worsen or improve (pejoration or amelioration), or it can change due to association (metaphor or metonymy). Semantic narrowing accounts for differences between cognates such as English "deer" and German "Tier [animal]." The English cognate "deer" once meant "animal" (cf. Old English "deor [animal]"), but its meaning narrowed to refer to a specific type of animal—namely, one with antlers. If English-speaking L2 learners of German are informed about this connection, because the word "deer" is already stored in their long-term memory, in theory they should be able to create a form–meaning L1–L2 connection. Pejoration (i.e., the worsening of meaning) accounts for cognate differences such as English "silly" and German "selig [holy]," where "silly" became a pejorative in English, but the meaning of sacredness is still retained in German. Change by association accounts for why English has "bead" and German has "Gebet [prayer]," where the association of praying and rosary beads led to a change in meaning from "pray" to "bead" in English (cf. Old English "gebed [prayer]" and "biddan [to pray]").

METHODOLOGY

Demographic

Two sections of third-semester German at an American midwestern university took part in this study, divided into two learning conditions: an intentional condition ($n = 18$) and an incidental condition ($n = 17$).[9] To account for the instructor as a potential confound, the same instructor was used to teach both sections. English was the L1 for an average of 69% of the learners (intentional = 61%, incidental = 76%). Thirty-one percent of the learners had a non-English L1, but all nonnative speakers were proficient speakers of English based on their Test of English as a Foreign Language (TOEFL) and English proficiency scores. A non-English L1 background (intentional = Russian 11%, Mandarin 11%, Spanish 11%, Portuguese 6%; incidental = Mandarin 12%, Spanish 6%, Hindi 6%) made it possible to explore the effects of using declarative knowledge of English to aid L2 (or L3) learners of German for both English L1 and non-English L1 speakers. Other than rudimentary knowledge of Spanish and Italian, no learners reported having any L2 knowledge of a language other than English and German. No learners in either learning condition reported knowledge of a Germanic language beyond English and German. As for the mean previous exposure of German, the incidental condition had learned German for slightly longer ($M = 3.5$ years) than the intentional condition ($M = 2.0$ years).

Assessment

Over the course of a 16-week semester, learners in both learning conditions completed a pretest (during Week 1), posttest (during Week 9), and delayed posttest (during Week 16). In an isolated translation task, learners were asked to provide an L1 translation for a list of 126 German words. Although nonnative speakers of English were given the option to provide a translation in their L1, all learners provided translations in English. Given the nature of the task, it tested receptive as opposed to productive vocabulary knowledge, it focused specifically on the "meaning" category of vocabulary knowledge (Nation, 2013), and it tested vocabulary size as opposed to vocabulary depth. The order of the 126 words was counterbalanced with three possible orders, and the tests were administered on paper during regular class time.

Of the 126 words, 50% ($n = 63$) were target cognates and 50% ($n = 63$) were distractors. Of the 63 target cognates, 42 were cognates that both learning conditions would encounter during their six 20-minute training sessions. The remaining 21 target cognates were words that learners would not encounter in the training sessions. Dividing the cognates into words that learners would encounter and words they would not encounter made it possible to test the impact of the learning conditions on both cognate acquisition and cognate predictability.[10] Moreover, having a minimum of 21 test items per RQ for each participant is in line with the recommendations of Loewen & Hui (2021), where, in the face of a small sample size, researchers are advised to increase the number of test items in order to increase the reliability of the findings. Of the 42 words that learners would encounter in the training sessions, 50% ($n = 21$) were cognates that were chosen because they are less recognizable due to historical sound changes, and 50% ($n = 21$) were cognates that are less recognizable because of historical semantic changes. Dividing the stimuli in this way also made it possible to examine the effect of the two different types of historical changes (sound and semantic). Table 1 provides a quantitative summary of the words on the tests, Table 2 provides a qualitative breakdown of the 21 target cognates affected by semantic changes, and Table 3 provides a

TABLE 1
Summary of Words on the Test

Word Type	*n*
Distractors	63
Cognates	63
Encountered	42
Unencountered	21

Note. "Encountered" refers to cognate items included in training sessions; "unencountered" words were not included in training sessions.

qualitative breakdown of the 42 target cognates that had undergone sound changes. A reliability analysis found a Cronbach's alpha of $\alpha = .818$ for the responses on the target cognates ($n = 63$), suggesting a high confidence level for the reliability of the testing instrument.

Coding

Responses on the isolated translation task were graded on a linear scale between 0 and 1. Fully correct answers received a score of 1 and incorrect answers received a score of 0. Answers that were partially correct received a score between 0 and 1. This linear grading system was necessary because, given the nature of the stimuli, learners could correctly identify the English cognate, but due to a lack of knowledge of the historical semantic changes, fail to provide a correct translation. For instance, the German word "Bein" is cognate with English "bone," but in German, "Bein" means "leg" not "bone." Therefore, learners who wrote "bone" as a translation received a score of .5 for correctly identifying the English cognate but did not receive a score of 1 for failing to identify the contemporary meaning. Learners who correctly identified the meaning, but chose the wrong part of speech, received a score of .75 (e.g., "offen [open]" erroneously translated as "to open").[11]

Intervention

After completion of the pretest in Week 1, between Weeks 2 and 7, learners in the intentional condition received six 20-minute training sessions on historical linguistics as it relates to the learning of English–German cognates (see Appendix A). During this time, learners received explicit instruction on two historical sound changes: Second Germanic Sound Shift (Salmons, 2012) and Ingvaeonic Palatalization (Lass, 1994). Because learners were not linguists, for pedagogical reasons, the rules were simplified and linguistic terminology (such as "fricatives," "affricates," etc.) was avoided. Ingvaeonic Palatalization was therefore explained to the intentional condition in the following way: "Words which have been around for at least 1,500 years, which start with a 'k' sound in German, changed their initial sound to 'ch' after certain vowels in English." Similarly, parts of the Second Germanic Sound Shift were explained in the following way: "Words which start with 'p' in English changed their beginning sound to 'pf' in German." The intentional condition also received explicit instruction on general changes in meaning (e.g., narrowing, broadening) insofar as these helped to build a connection between the English and German cognates. Similarly, during Weeks 2–7, the incidental group spent six 20-minute sessions carrying out various reading and communicative-based activities (see Appendix B). Learners in the incidental condition encountered the target cognates but did not receive any explicit historical instruction. The training sessions were conducted in the classroom during regular class time.

RESULTS

Acquisition Accuracy of Encountered Cognates

Overall Scores. The descriptive statistics in Table 4 indicate that the intentional condition outperformed the incidental condition in the learning of the 42 encountered cognates. On average, learners in the intentional condition learned an additional 19 words from pretest ($M = .12$, $SD = .33$) to posttest ($M = .57$, $SD = .49$). In contrast, learners in the incidental condition learned only one additional word from pretest ($M = .11$, $SD = .31$) to posttest ($M = .15$, $SD = .11$).

To explore this difference further, a series of linear mixed models (LMMs) were run in SPSS 26. An $\alpha = .05$ was used as the criterion for significance, and Cohen's d was used to measure effect size using the benchmarks of Plonsky & Oswald (2014): small ($d = .40$), medium ($d = .70$), large ($d = 1.0$). All pairwise comparisons were performed with Sidak correction. The random effects structure was the same for all models—that is, learner and word were run as random intercepts. All reported confidence intervals are at 95% confidence.

An initial omnibus model was run, with translation accuracy as the dependent variable and group, time, and learner-L1 as fixed factors. Group had two levels (intentional condition,

TABLE 2
Target Cognates Affected by Semantic Changes

Cognate	Semantic Relationship
1. weh [pain]	cognate "woe"
2. sterben [to die]	cognate "to starve"—semantic narrowing in English
3. Weib [woman (pej)]	cognate "wife" (OE "wīf") used to mean "woman"
4. versehren [to injure]	cognate "sore;" related to German "sehr [very]," which used to mean "pain"
5. Zimmer [room]	cognate "timber"—change by association in German
6. Vogel [bird]	cognate "fowl" (OE "fugol")—semantic narrowing in English
7. Gebet [prayer]	cognate "bead"—change by association
8. beten [to pray]	cognate "bead" (same as "Gebet")
9. Zwilling [twin]	cognate "two;" German "zw" is English "tw," e.g., "zwischen [between]"
10. Knecht [servant]	cognate "knight" (OE "cniht")—amelioration in English
11. Tier [animal]	cognate "deer" (OE "deor")—semantic narrowing in English
12. satt [full]	cognate "sad" originally meant "full" as in "satisfy"
13. selig [holy]	cognate "silly"—pejoration in English
14. Waren [goods]	cognate "–ware," as in "silverware," "hardware," and "warehouse"
15. Burg [fortress]	cognate "–burg(h)," as in Edinburgh (people used to live in a "Burg")
16. Bürger [citizen]	cognate "–burg(h);" people who lived in a Burg were "Bürger" (lit. "of the Burg")
17. Zaun [fence]	cognate "town" (OE "tūn"); original meaning was "enclosed space"
18. Bein [leg]	cognate "bone"—specialization in German
19. reißen [to rip]	cognate "to write" (OE "wrītan"); people used to rip or carve into wood to "write" something
20. Urlaub [holiday]	cognate "to allow;" it was necessary to ask permission to take "leave"
21. wissen [to know]	cognate "wit" (OE "witan [to know]"); relict "to have your wits about you"

Note. OE = Old English.

incidental condition), time had three levels (pretest, posttest, delayed posttest), learner-L1 had two levels (English, non-English), and translation accuracy was continuous. All possible interactions were included in the model: two-way interactions of Group × Time, Group × Learner-L1, and Time × Learner-L1; and a three-way interaction of Group × Time × Learner-L1.

The omnibus model demonstrated a significant effect of group, $F(1,4398) = 27.656$, $p < .001$, and time, $F(2,4398) = 138.307$, $p < .001$, but not learner-L1, $F(1,4398) = .862$, $p = .350$. The effect of group was due to a significantly higher translation accuracy in the intentional condition ($M = .40$, $SD = .48$) compared to the incidental condition ($M = .14$, $SD = .33$), with an effect size of $d = .63$, 95% CI [.05, 1.31]. Post hoc pairwise comparisons indicated that the effect of time was due to significant differences in translation accuracy in the pretest ($M = .12$, $SD = .31$) compared to the posttest ($M = .38$, $SD = .48$) and delayed posttest ($M = .36$, $SD = .48$). However, the effect size from pretest to posttest ($d = .64$, 95% CI [.16, 1.10]) and pretest to delayed posttest ($d = .59$, 95% CI [.12, 1.10]) suggests that either the effect was small or there was some interference of group. As for the interaction between the three fixed factors, there was only a significant effect of Group × Time, $F(2,4398) = 88.756$, $p < .001$. The omnibus model therefore suggests that the historical instruction learners in the intentional condition received on sound and semantic changes had a significantly greater effect on the learning of the 42 cognates than the intervention the incidental condition received.[12]

To confirm the effect of time in the two learning conditions, two separate LMMs were run: one using translation accuracy in the intentional condition and one using translation accuracy in the incidental condition. The model on the intentional condition confirmed that there was a significant effect of time, $F(2,2262) = 317.904$, $p < .001$ from pretest ($M = .12$, $SD = .33$) to posttest ($M = .57$, $SD = .49$), with an effect size of $d = 1.0$, 95%

TABLE 3
Target Cognates Affected by Sound Changes[a]

Change	Encountered Cognates	Unencountered Cognates
	Ingvaeonic Palatalization	
"k" became "ch" before front vowels "i," "ä," and "e"	*Kinn* > chin	*Krücke* > crutch
	Käfer > chafer (type of beetle)	*strecken* > stretch
	Kerl > cherl (archaic word for man)	*kauen* > chew
	Second Germanic Sound Shift	
"p" became "pf" at the beginning of a word	pipe > *Pfeife*	penny > *Pfennig*
	pan > *Pfanne*	pole > *Pfahl*
	pound > *Pfund*	pepper > *Pfeffer*
"p" became "pf" between two vowels	to tap > *zapfen*	to hop > *hüpfen*
	copper > *Kupfer*	to stamp > *stampfen*
	drop (as in eye drops) > *Tropfen*	apple > *Apfel*
"p" became "f" before "n," "m," and "l"	open > *offen*	grip > *Griff*
	weapon > *Waffe*	sharp > *scharf*
	ripe > *reif*	to slurp > *schlürfen*
"t" became German "z," pronounced [t͡s], at the beginning of a word, and sometimes before a consonant	tongue > *Zunge*	twig > *Zweig*
	tin > *Zinn*	wart > *Warze*
	toe > *Zeh*	to fart > *furzen*
"t" became "s" at the beginning of a word and between two vowels	to let > *lassen*	kettle > *Kessel*
	hate > *Hass*	to sweat > *schweißen*
	foot > *Fuß*	nut > *Nuss*
"th" became "d" at the beginning of a word and between two vowels	thing > *Ding*	thorn > *Dorn*
	thirst > *Durst*	feather > *Feder*
	these > *diese*	thistle > *Dissel*

[a]The author acknowledges that some of these words are not necessarily frequent, such as "Zweig [twig]." However, the target words were chosen not because of frequency but because they are cognates.
Note. German cognates are italicized.

TABLE 4
Knowledge of Encountered Cognates (Descriptive Statistics)

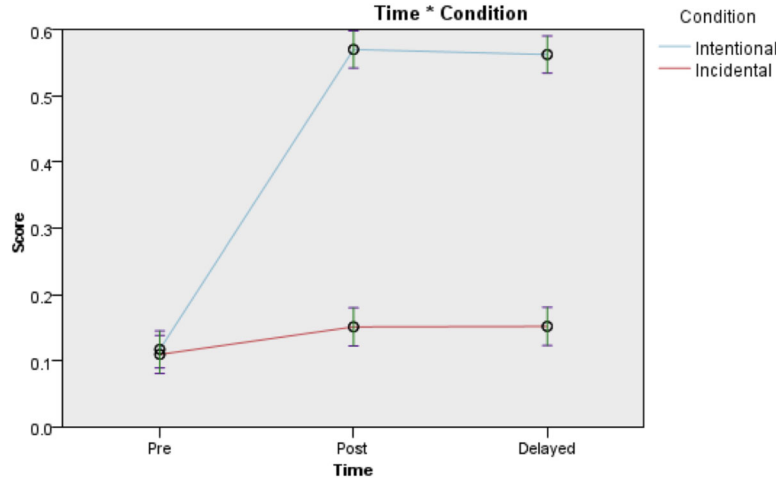
	Pretest			Posttest			Delayed Posttest		
Condition	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Intentional	89/756	.12	.33	431/756	.57	.49	425/756	.56	.48
Incidental	79/714	.11	.31	108/714	.15	.35	108/714	.15	.36

Note. The *n*, which refers to the number of correct answers versus the total number of questions, is out of 756 for the intentional condition but 714 for the incidental condition because there were 18 learners in the intentional condition (18 learners × 42 cognates = 756) and 17 in the incidental condition (17 learners × 42 cognates = 714). Mean (*M*) and standard deviation (*SD*) are for translation accuracy score (0–1).

CI [.38, 1.80], and from pretest ($M = .12$, $SD = .33$) to delayed posttest ($M = .56$, $SD = .48$), with an effect size of $d = 1.1$, 95% CI [.37, 1.80], but not from posttest ($M = .57$, $SD = .49$) to delayed posttest ($M = .56$, $SD = .48$).

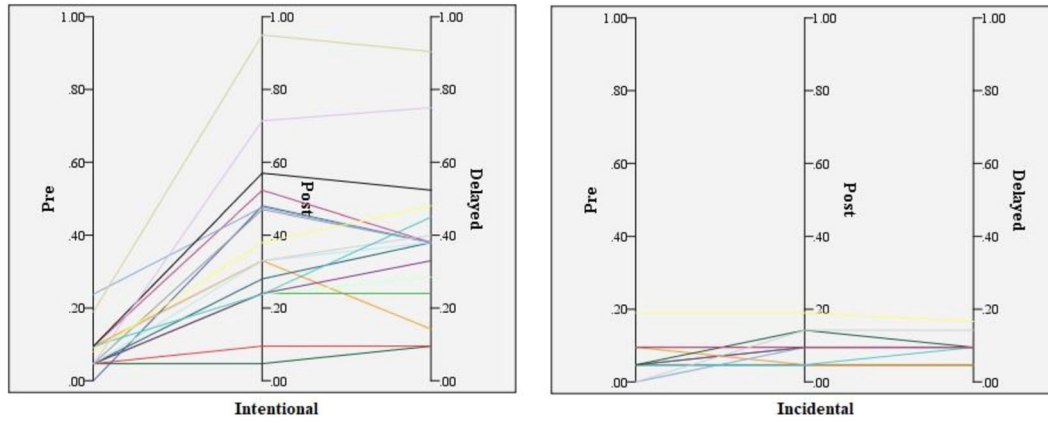
These results therefore confirm that the instruction the intentional condition received had a significant effect on the learning of cognates. Although the LMM run on the incidental condition also found a significant difference in pretest scores ($M = .11$, $SD = .31$) compared to posttest ($M = .15$, $SD = .11$) and delayed posttest scores ($M = .15$, $SD = .36$), the effect sizes were lower than the minimal threshold: from pretest to posttest, $d = .17$, 95% CI [−.50, .85], and from pretest to delayed posttest, $d = .12$, 95% CI [−.55, .79], suggesting that the intervention the intentional condition received had little effect on cognate acquisition. The mean differences in Figure 1 and the individual differences in Figure 2 illustrate this contrast. In the parallel coordinate plot in Figure 2, each line represents the average

FIGURE 1
Knowledge of Encountered Cognates by Condition From Pretest to Delayed Posttest
[Color figure can be viewed at wileyonlinelibrary.com]



Note. The whiskers represent the 95% confidence intervals for the mean scores.

FIGURE 2
Parallel Coordinate Plot of Individual Differences for Translation Accuracy of Encountered Cognates From Pretest to Delayed Posttest
[Color figure can be viewed at wileyonlinelibrary.com]



Note. The parallel coordinate plot was created in line with the recommendations of Larson–Hall (2017), where L2 researchers are encouraged to use data accountable graphics to show within-group individual differences.

mean score of each individual learner, which clearly shows that while each learner in the intentional condition increased their score from the pretest to posttest, this was not true for all learners in the incidental condition. Moreover, the growth or increase in translation accuracy was smaller in the incidental condition compared to the intentional condition. Therefore, the significance of group in the omnibus model, comparisons of effect sizes in the separate models, and mean and individual differences show that the intentional condition outperformed the incidental condition with respect to translation accuracy.

The nonsignificance of learner-L1 in both the intentional $F(1,2262) = 1.326$, $p = .250$, and incidental condition $F(1,2136) = .030$, $p = .861$, indicates that the learners' L1 had no impact on the effect of the two interventions. Learners with a non-English L1 performed equally as well as learners whose L1 was English. This finding therefore suggests that taking advantage of the historical relationship between English and German

TABLE 5
Knowledge of Encountered Cognates Affected by Semantic Changes From Pretest to Delayed Posttest

	Pretest			Posttest			Delayed Posttest		
Condition	*n*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Intentional	66/378	.17	.38	229/378	.60	.48	230/378	.61	.47
Incidental	60/357	.19	.37	82/357	.23	.41	77/357	.22	.41

Note. The *n*, which refers to the number of correct answers versus the total number of questions, is out of 378 for the intentional condition but 357 for the incidental condition because there were 18 learners in the intentional condition (18 learners × 21 cognates = 378) and 17 in the incidental condition (17 learners × 21 cognates = 357). Mean (*M*) and standard deviation (*SD*) are for translation accuracy score (0–1).

when learning cognates is effective in an English-speaking classroom, so long as learners are proficient speakers of English.

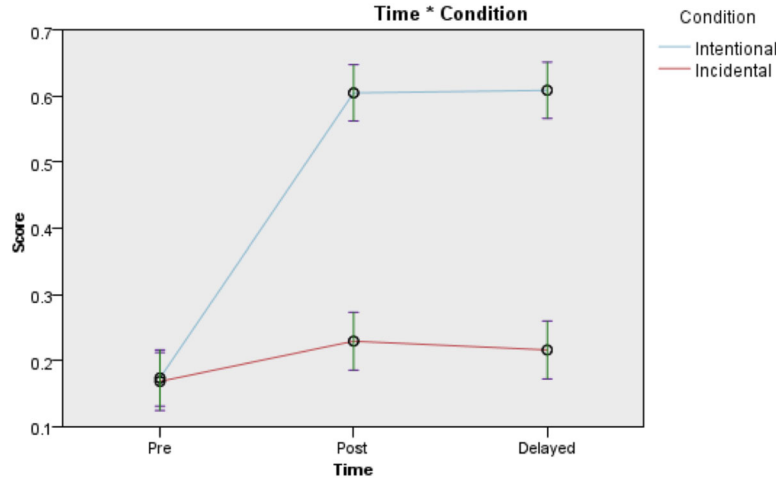
Acquisition of Cognates Affected by Semantic Changes. The descriptive statistics in Table 5 indicate that the intentional condition outperformed the incidental condition for the learning of cognates affected by semantic changes. Each learner in the intentional condition learned an average of nine words from pretest ($M = .17$, $SD = .38$) to posttest ($M = .60$, $SD = .48$) compared to no more than one word from pretest ($M = .19$, $SD = .37$) to posttest ($M = .23$, $SD = .41$) in the incidental condition. To examine the specific effect of receiving historical semantic instruction further, an LMM was run on the translation scores provided for the 21 cognates affected by semantic changes. Group, time, and learner-L1 were run as fixed factors, and translation accuracy was run as the dependent variable.

The model found a significant effect of group, $F(1,2193) = 25.293$, $p < .001$, and time, $F(2,2193) = 64.210$, $p < .001$, as well as a significant interaction between the two, $F(2,2193) = 31.441$, $p < .001$. As in the previous model, learner-L1 was not statistically significant ($p = .067$). The effect of group was due to significant differences in the translation accuracy of the cognates affected by semantic changes between the intentional ($M = .48$, $SD = .50$) and incidental condition ($M = .19$, $SD = .40$), with an effect size of $d = .64$, 95% CI [.16, 1.12], suggesting that the semantic explanations the intentional condition received had a greater effect on the learning of German than the incidental learning activities. A summary of the mean differences is provided in Figure 3.

Although the intentional condition outperformed the incidental condition, learners in both learning conditions made theoretically interesting errors. First, even though the intentional condition improved from posttest to delayed posttest, many learners in the delayed posttest provided only the English cognate as a translation for the German word, as opposed to the modern contemporary translation. For instance, three learners correctly translated "Weib" as "woman" in the posttest but put only the cognate "wife" as a translation in the delayed posttest. Four learners correctly translated "Zaun" as "fence" in the posttest but wrote "town" as a translation in the delayed posttest. These errors reveal that receiving a semantic explanation on the relationship between English and German cognates allowed learners to create an English–German form–meaning link. Even 2 months after receiving the intentional instruction, the English–German connection was retained. Given that the intentional condition outperformed the incidental condition in the learning of these cognates, it is clear that associating previous knowledge with novel information can aid L2 vocabulary learning.

The second error concerned the incidental condition. Based on their translations, it was evident that this learning condition was more prone to over- and undergeneralization of word meaning. For instance, by the posttest, four learners wrote "pet" as a translation of "Tier [animal]," likely due to the influence of the compound noun "Haustier [pet]" ("Haus [house]" + "Tier [pet]"). Three learners wrote "mayor" as a translation of "Bürger [citizen]," likely due to the influence of the compound noun "Bürgermeister [mayor]" ("Bürger [citizen]" + "Meister [master]"). Several different semantically related translations appeared for "Waffe [weapon]," such as "sword," "army," "military," and "group." One learner also translated "Zimmer [room]" as "class," likely due to the influence of the compound "Klassenzimmer [classroom]" ("Klasse [class]" + "n" + "Zimmer [room]"). These responses suggest that, under incidental learning conditions, learners are

FIGURE 3
Knowledge of Encountered Cognates Affected by Semantic Changes From Pretest to Delayed Posttest [Color figure can be viewed at wileyonlinelibrary.com]



Note. The whiskers represent the 95% confidence intervals for the mean scores.

TABLE 6
Knowledge of Encountered Cognates Affected by Sound Changes (Descriptive Statistics)

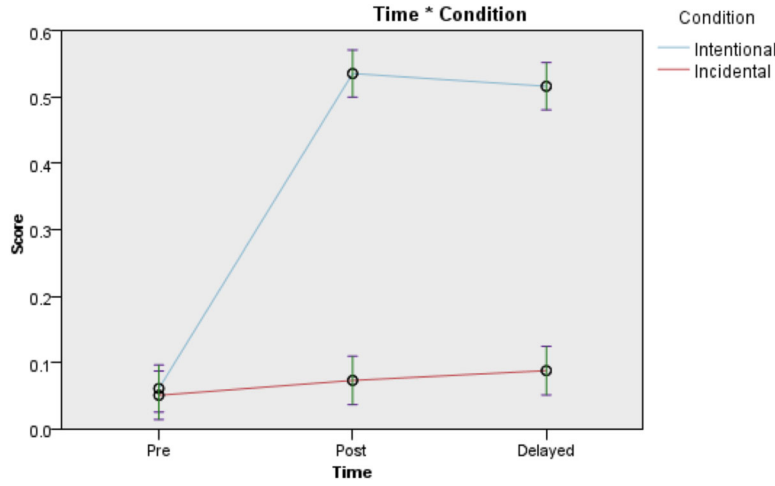
	Pretest			Posttest			Delayed Posttest		
Condition	*N*	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
Intentional	23/378	.06	.25	203/378	.54	.49	195/378	.52	.50
Incidental	18/357	.05	.22	26/357	.07	.26	32/357	.09	.28

Note. The *n*, which refers to the number of correct answers versus the total number of questions, is out of 378 for the intentional condition but 357 for the incidental condition because there were 18 learners in the intentional condition (18 learners × 21 cognates = 378) and 17 in the incidental condition (17 learners × 21 cognates = 357). Mean (*M*) and standard deviation (*SD*) are for translation accuracy score (0–1).

not always able to correctly infer word meaning. When learners infer, they can arrive at a meaning that is in the correct semantic field contextually (e.g., "Zunge [tongue]" translated as "tooth" and "Bein [leg]" translated as "knee" and "arm"), but their inferred meaning may be over- or undergeneralized. Moreover, when learners encounter words in compound structures under incidental or implicit conditions, because of their lack of metalinguistic awareness, they often fail to recognize morpheme boundaries (i.e., "Haustier" = "Haus" + "Tier") and consequently generalize the meaning of the compound (e.g., "Haustier [pet]") to the noncompound word form (e.g., "Tier [animal]" erroneously translated as "pet").

Acquisition of Cognates Affected by Sound Changes. As the descriptive statistics in Table 6 indicate, the intentional condition provided more correct translations for the cognates affected by sound changes than the incidental condition. The intentional condition learned an average of nine cognates from pretest ($M = .06$, $SD = .25$) to posttest ($M = .54$, $SD = .49$), compared to one word or fewer from pretest ($M = .05$, $SD = .22$) to posttest ($M = .07$, $SD = .26$) in the incidental condition. To examine this difference further, an LMM was run on the translation accuracy of the 21 cognates affected by historical sound changes. The LMM found a significant effect of group, $F(1,2193) = 25.736$, $p < .001$, and time, $F(2,2193) = 83.147$, $p < .001$, as well as a significant effect of Group × Time, $F(2, 2193) = 68.354$, $p < .001$. Learner-L1, however, was not statistically significant ($p = .996$). As Figure 4 illustrates, the effect of group was due to significant differences in translation accuracy between the incidental ($M = .05$, $SD = .25$) and intentional condition ($M = .33$, $SD = .48$),

FIGURE 4
Knowledge of Encountered Cognates Affected by Sound Changes From Pretest to Delayed Posttest
[Color figure can be viewed at wileyonlinelibrary.com]



Note. The whiskers represent the 95% confidence intervals for the mean scores.

suggesting that declarative knowledge of the Second Germanic Sound Shift and Ingvaeonic Palatalization had a significantly greater effect than incidental vocabulary learning on the learning of the cognates ($d = .73$, 95% CI [.05, 1.40]). According to post hoc pairwise comparisons, the effect of time was due to significant differences in the pretest ($M = .05$, $SD = .23$) when compared to the posttest ($M = .33$, $SD = .46$) and delayed posttest ($M = .29$, $SD = .45$). The significant effect of Group × Time suggests that the learning conditions performed differently over time, with a significantly larger improvement in the intentional condition than in the incidental condition (see Figure 4).

Two follow-up models were run to confirm these effects. The first model run on translation accuracy in the intentional condition found a significant effect of time, $F(2,1128) = 177.143$, $p < .001$. Post hoc pairwise comparisons indicated that the difference from pretest ($M = .06$, $SD = .25$) to posttest ($M = .54$, $SD = .49$) was significant ($p < .001$), with an effect size of $d = 1.2$, 95% CI [.52, 2.00]. The difference from pretest ($M = .06$, $SD = .25$) to delayed posttest ($M = .52$, $SD = .49$) was significant ($p < .001$), with an effect size of $d = 1.2$, 95% CI [.47, 1.90], but the difference in posttest ($M = .54$, $SD = .49$) and delayed posttest ($M = .52$, $SD = .49$) was not statistically significant, $p = .471$, $d = .04$, 95% CI [−.69, .61]. In contrast, the second model on the incidental condition found no significant effect of time, $F(2,1065) = 545$, $p = .580$. Therefore, the combined effects of the different LMMs indicate that although both learning conditions acquired the meaning of some German cognates over the course of the 16-week semester, improvement was more gradual in the incidental condition than the intentional condition.

Prediction Accuracy of Unencountered Cognates

As for the unencountered cognates, the descriptive statistics (see Table 7) show that the intentional condition outperformed the incidental condition, with learners in the intentional condition correctly predicting the meaning of, on average, an additional six unencountered cognates from pretest ($M = .07$, $SD = .26$) to posttest ($M = .38$, $SD = .49$). The cognates whose meaning were most frequently correctly predicted were "Pfennig [penny]," "Nuss [nut]," "Griff [grip/handle]," "hüpfen [to hop]," "kauen [to chew]," "Apfel [apple]," and "stampfen [to stamp]"; the latter two whose meanings were correctly predicted by most learners in both conditions in the pretest.

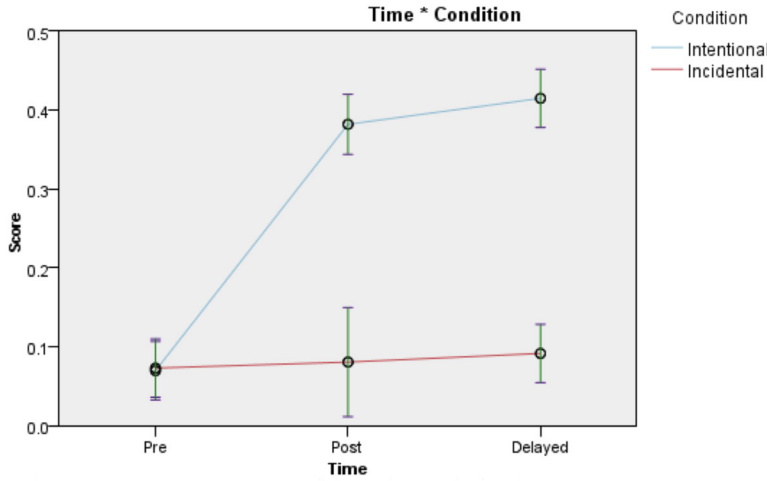
To examine differences in prediction accuracy of the 21 unencountered cognates between the two conditions further, an omnibus model (LMM) was run with prediction accuracy as the dependent variable and group, time, and learner-L1 run as fixed factors. As in the previous models, learner and word were run as random intercepts, and all possible interactions were included in the model. The model found a significant

TABLE 7
Knowledge of Unencountered Cognates (Descriptive Statistics)

	Pretest			Posttest			Delayed Posttest		
Condition	*n*	*M*	*SD*	*N*	*M*	*SD*	*n*	*M*	*SD*
Intentional	27/378	.07	.26	136/378	.38	.49	157/378	.42	.49
Incidental	27/357	.07	.26	31/357	.09	.28	32/357	.09	.29

Note. The *n*, which refers to the number of correct answers versus the total number of questions, is out of 378 for the intentional condition but 357 for the incidental condition because there were 18 learners in the intentional condition (18 learners × 21 unencountered cognates = 378) and 17 in the incidental condition (17 learners × 21 unencountered cognates = 357). *M* and *SD* are for translation accuracy score (0–1).

FIGURE 5
Knowledge of Unencountered Cognates From Pretest to Delayed Posttest
[Color figure can be viewed at wileyonlinelibrary.com]

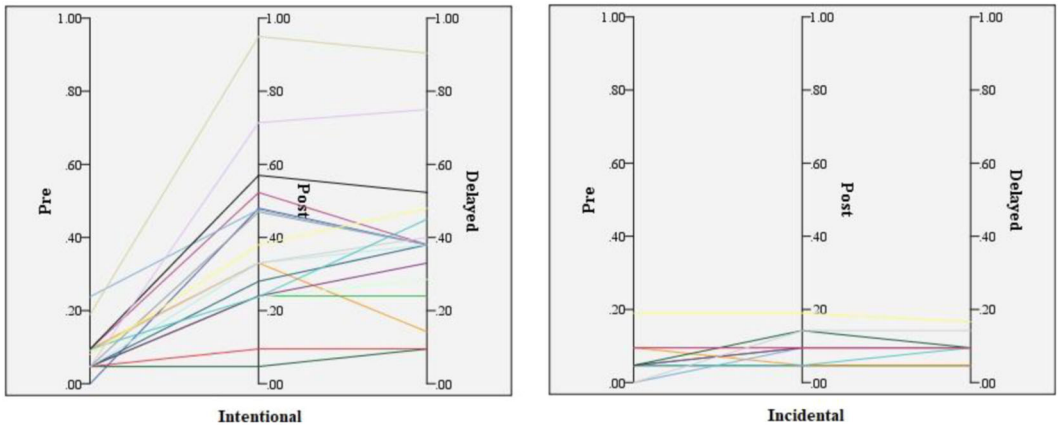


Note. The whiskers represent the 95% confidence intervals for the mean scores.

effect of time, $F(2,2193) = 15.372$, $p < .001$, and group, $F(2,2193) = 41.890$, $p < .001$, a significant interaction of Group × Time, $F(2,2193) = 18.513$, $p < .001$, but no significant effect of learner-L1 $F(2,2193) = .010$, $p = .994$. The effect of group was due to significant differences in prediction accuracy between the incidental ($M = .10$, $SD = .29$) and intentional condition ($M = .29$, $SD = .50$), with an effect size of $d = .46$, 95% CI [.21, 1.20]. Post hoc pairwise comparisons indicated that the effect of time was due to significant differences in prediction accuracy in the pretest ($M = .05$, $SD = .25$) when compared to the posttest ($M = .24$, $SD = .44$), with an effect size of $d = .53$, 95% CI [.05, 1.10], and the delayed posttest ($M = .24$, $SD = .43$), with an effect size of $d = .44$, 95% CI [.04, .91]. As Figure 5 illustrates, the interaction of Group × Time indicates that the prediction accuracy in the intentional condition significantly improved over time. This improvement is also clearly shown in the individual changes in Figure 6.

To confirm these results, two follow-up models were run on the prediction data from the two groups. The results confirmed time as a significant factor in the intentional condition, $F(2,1128) = 71.033$, $p < .001$, but not in the incidental condition, $F(1,1065) = 1.571$, $p = .340$, indicating that incidental learning did not have a significant effect on the ability to correctly predict the meaning of English–German cognates, whereas the explicit instruction the intentional condition received did. The prediction accuracy in the intentional condition was stable from posttest to delayed posttest, suggesting that the declarative knowledge learners in this condition had obtained remained with them approximately 2 months after the intervention. Post hoc pairwise comparisons indicated that the pretest ($M = .07$,

FIGURE 6
Parallel Coordinate Plot of Individual Differences in Prediction Accuracy of Unencountered Cognates From Pretest to Delayed Posttest [Color figure can be viewed at wileyonlinelibrary.com]



SD = .26) compared to the posttest (*M* = .38, *SD* = .49) or delayed posttest (*M* = .42, *SD* = .49) was significant. The effect size from pretest to posttest was *d* = .79, 95% CI [.11, 1.50], and from posttest to delayed posttest was *d* = .89, 95% CI [.21, 1.60].

Whereas the results suggest that learners in the intentional condition relied upon their newly acquired declarative knowledge to predict the meaning of the unencountered cognates, the responses from the incidental condition suggest that the incidental group was left to erroneously infer cognate meaning based on the similarity of the words in English. For instance, in the incidental condition, "Kessel [kettle]" was often translated as "castle," "schweißen [to sweat]" was often translated as "to swipe" or "switch," and "Dorn [thorn]" was often translated as "dorm." Similar guessing strategies were also present for cognates the incidental condition had encountered. For instance, in the incidental condition's posttest, "Bürger [citizen]" was translated as "burger" by five learners, "offen [open]" was translated as "often" by ten learners, and "Kinn [chin]" was translated as "family/child/kids" by four learners. In contrast, although in the intentional condition, two learners translated "Waffe" as "waffle" in the pretest and seven learners translated "Krücke [crutch]" as "crook" in the pretest, these incorrect translations were corrected by the posttest. Qualitative evidence also suggests that the incidental condition was more sensitive to "synformy" or "synforms" (i.e., the similarity of word forms; Laufer, 1988) than the intentional condition. This was true for both encountered (e.g., "reißen [to rip]" confused with "reisen [to travel]," "Gebet [prayer]" confused with "Gebiet [area]") and unencountered cognates (e.g., "Zweig [twig]" confused with "Zwerg [dwarf]," "Griff [grip/handle]" confused with "Giraffe [giraffe]" and "Angriff [attack]").

DISCUSSION

The present study set out to explore whether knowledge of language history can be beneficial for learners when learning English–German cognates. The first RQ investigated whether there was a significant difference in the number of cognates learned in the two learning conditions, as measured by a translation task. Results indicate that the intentional condition significantly outperformed the incidental condition in translation accuracy. One logical interpretation of this result is that the historical instruction that learners in the intentional condition received had a significant effect on the learning of the cognates due to the degree of elaboration. In other words, the historical explanations provided learners in the intentional condition with a method to associate novel L2 cognates with their already existing English lexicon, regardless of whether the learners were L1 speakers of English or not. The effect of elaboration (Craik & Tulving, 1975; Craik & Watkins, 1973) has been demonstrated for decades in work on human memory and learning (Bolger & Zapata, 2011; Bower & Clark, 1969; Craik & Lockhart, 1972; Kirsch, 2012; Laufer & Hulstijn, 2001; McNamara & Scott, 2001; Prince, 2012). However, the somewhat novel contribution of the findings from the present study is that historical narratives, such as being cognizant of the etymological association between L1 and L2

cognates (specifically English–German cognates), and not just fictional narratives (as in Bower & Clark, 1969) may significantly aid the vocabulary acquisition process in the L2 classroom.[13] While scholars have called for historical instruction in the L2 German classroom (Horsford, 1987; Lightfoot, 2007; Wolff, 1993), prior to the present study, no empirical studies had investigated its effect on the learning of German L2 vocabulary. Some studies had been carried out on other language pairings such as English–French (Arteaga & Herschensohn, 1995), but no studies had tested the joint effects of receiving explicit instruction on both sound and semantic changes.

In line with previous observations (Carpenter et al., 2012; Hulstijn, 1992), learners in the incidental group were more prone to inference errors. While it was clear from the translations on the posttest and delayed posttest that learners in the incidental condition made attempts to infer word meaning, their inferences often resulted in over- and undergeneralization. This was particularly prominent with compound nouns, where a lack of metalinguistic awareness of morpheme boundaries appears to have resulted in incorrect generalizations (e.g., "Tier [animal]" erroneously translated as "pet" due to "Haustier [pet]"). After all, metalinguistic awareness is thought to account for individual differences in incidental vocabulary acquisition (Nagy, 2007). Learners in the incidental condition were also more sensitive to confusion of synforms (Laufer, 1988), which may be a consequence of meaning-focused instruction in general. Interestingly, however, many incorrect translations in the intentional condition in the pretest (e.g., "Krücke [crutch]" translated as "crook") were not repeated by the posttest. In contrast, errors in the incidental condition (e.g., "Bürger [citizen]" was translated as "burger") were repeated across all three tests. This finding is in line with research that suggests that incorrect inferences are repeated unless corrected (Carpenter et al., 2012).

The second RQ set out to examine the effect of historical instruction on the ability to correctly predict the meaning of cognates. As was anticipated, the intentional condition significantly outperformed the incidental condition, most likely due to the acquired declarative knowledge of the historical sound changes. Even if learners had somehow encountered the "unencountered" cognates outside of the classroom, the probability of encountering these words was the same in both learning conditions, yet the incidental condition performed significantly worse. Since the 21 unencountered target cognates were affected by either Ingvaeonic Palatalization or the Second Germanic Sound Shift, learners in the intentional condition were able to use this knowledge to deduce many cognate meanings. Because acquiring a language is more than just imitation, repetition, and rote memorization of input, it seems only fair to provide L2 learners with a tool kit (when available) to allow them to correctly infer meaning themselves. Work on reading shows that outside knowledge is crucial in L2 reading comprehension (Anderson, 2004), and the present study shows that similar outside knowledge (i.e., knowledge of the history of the language) can be just as effective when learning L2 vocabulary.

There are, however, two notable caveats that need to be acknowledged in interpreting and generalizing the results from this study. First, even though the intentional group outperformed the incidental group, there was a bias toward explicit testing. Although the incidental group did not receive any explicit instruction on the cognates, the translation tests called for explicit knowledge. Since research shows that assessments that match the learning conditions yield higher performance (Baddeley, 1990), the method of testing may have impacted the results. However, that said, the majority of assessments used in the North American L2 classroom call upon explicit knowledge, even if the instruction or learning conditions are predominantly incidental or implicit (Schmitt, 2010).

Second, the results from this study should be contextualized within the bounds of the used sample size. Although the sample size was in line with the average size used in classroom-based L2 acquisition studies (Plonsky, 2013), as Loewen and Hui (2021) pointed out, a sample size of 19 or lower for each learning condition is insufficient to make strong claims about the broader efficacy of a given type of instruction. Although the reported mean differences, individual differences, statistical significance differences, and effect size measures from the present study point toward the potential benefits of explicit historical instruction when learning a language that is historically related to a language that learners are familiar with, the broad range for the reported confidence intervals supports Loewen and Hui's observation about small sample sizes. If a larger sample size had been used and the same differences between groups were found, the confidence intervals would have been tighter and thus more reliable, which would have put more confidence in the generalizability of the study's findings to the L2 classroom. At this stage, it is clear that further follow-up studies with larger sample sizes are necessary to confirm the broader

generalizability of these findings. Naturally, the used small sample size was a natural by-product of carrying out this study in a classroom-based setting, which, according to the American Council on the Teaching of Foreign Languages (ACTFL) guidelines (ACTFL, 2012), should not exceed a 15:1 student–teacher ratio. However, Loewen and Hui suggested a number of ways, such as multisite research collaboration, that instructed L2 acquisition researchers can use to deal with low sample size constraints without affecting the student–teacher ratio. Nevertheless, the findings from the present study point toward potential benefits of the use of historical instruction in the L2 classroom and have laid the foundation and paved the way for future research into scientific inquiry of this kind, especially in the context of instructed L2 vocabulary acquisition.

CONCLUSION

In the last few decades, L2 vocabulary research has placed great emphasis on incidental vocabulary learning. While incidental learning clearly has a place in the L2 classroom, incidental L2 vocabulary acquisition is slow and gradual. Even though learning vocabulary intentionally can be beneficial to learners, explicit learning of this kind is often discouraged in the L2 classroom. This study highlighted the potential benefits of explicit instruction on language history when learning languages that are historically related. While the content and goals of a language class are clearly different to those of a course in historical linguistics, historical tidbits seem to be useful to learners, when relevant, to create an instant connection between the novel stimulus (i.e., the German words) and the information already stored in their long-term memory (i.e., the English cognate counterparts).

The positive effects of this applied historical instruction may add a new dimension to the term applied historical linguistics, with applications to the L2 classroom.[14] If integrated appropriately and explained in terms that nonspecialists can understand, speakers of English or other Germanic languages may have an advantage when learning historically related languages. While the focus in this study was on English-speaking L2 learners of German, it is likely this type of instruction would be just as valuable for German-speaking L2 learners of English. Similarly, it is possible that English-speaking L2 learners of other Germanic languages (e.g., Norwegian) could also benefit from similar types of instruction, as they too were affected by changes such as the Second Germanic Sound Shift.[15] This type of instruction may also be appropriate for more distantly related language pairings. One could speculate that knowledge of the First Germanic Sound Shift (as opposed to the second) could be beneficial for learners of other Indo-European languages (e.g., Russian, Spanish). Instruction on language history and language contact could also be beneficial when learning nonhistorically related languages. While the focus in this study was on cognates, it is possible that similar types of instruction could be useful when learning loanwords. In situations where high volumes of borrowing have taken place, such as the many English words borrowed into Japanese, loanwords have to be modified and accommodated appropriately according to the sounds available and the phonological constraints in the receiving language. If explicit attention is drawn to these patterns, they too could help learners create a connection between L1 and L2 vocabulary items, in turn, helping learners increase their vocabulary size. All in all, this study suggests that short bursts of historical instruction may have a place in the language classroom, at least when learning historically related languages.

NOTES

[1] In historical linguistics, an asterisk is used to refer to the reconstructed form of a word. A distinction is also made between loanwords and cognates whereas, in L2 acquisition research, cognates can also include borrowings (Peters & Webb, 2018). Since the role of historical linguistics is an important part of the methodology of the present study, the traditional historical definition of a cognate excluding loanwords is maintained.

[2] A stop consonant is a sound created by completely stopping the airflow (hence, "stop"). These include sounds such as /p, b, t, d, k, g/. In contrast, a fricative is a consonant which is produced with some obstruction, but the airflow is not completely stopped. Fricatives have a sizzling-like audible quality and include sounds such as the English "th" [θ/$\eth$] and "sh" [ʃ]. Interdental refers the place where the sound is produced, namely, between ("inter") the teeth ("dentes").

[3] Metalinguistic awareness refers to learners' explicit, declarative knowledge of language, or the phonological, morphological, lexical, syntactic, and pragmatic features of the L2.

[4] The labels "explicit" and "nonexplicit" could therefore also be used to describe the differences between the two learning conditions.

[5] The keyword method is an intentional learning strategy that involves establishing an acoustic link between a novel L2 item and the learners' L1. For instance, to learn the German adjective "lecker [tasty]," learners

may think of the image of someone licking something tasty. In so doing, learners create a connection between German "lecker" and English "lick."

[6] The bar above /t͡s/ refers to the fact that it is an affricate, which is a sound consisting of a stop and fricative. The "ch" sound in English in words like "church" is an example of an affricate since, phonetically speaking, it consists of the stop [t], as in "Tom," followed by a [ʃ], as in "she." The bar is not used in spelling, only in phonetic transcription.

[7] Assimilation describes the process whereby one sound becomes similar to a neighbouring sound. In the case of "mb" > "mm," the [b] fully assimilates to [m] because it became the same sound, that is, [m].

[8] Front vowels are those produced at the front of the mouth as opposed to back vowels, which are produced at the back of the mouth. For pedagogical simplicity, front vowels are "i," "ä," "ü," "ö," and "e," whereas back vowels are "u" and "o."

[9] The sample size was determined by the number of students who enrolled in the two sections. While post hoc power analyses revealed that the used sample size ($n = 35$) was appropriate on the basis of the effect sizes reported in the results section, the author acknowledges the controversy concerning the validity of such post hoc measures (cf. Loewen & Hui, 2021).

[10] Because the words on the posttest and delayed posttest were identical to those on the pretest, technically both learning conditions had been exposed to the 21 unencountered cognates. However, neither learning condition encountered the unencountered cognates in their six 20-minute sessions.

[11] The author acknowledges that this coding scheme creates a scale that is not truly an interval measure. In other words, there is no underlying continuum from 0 to 1. The fact that correct answers are coded 1, incorrect 0, and partially correct 0.5, does not mean that correct is twice as big as incorrect, even though the statistical analysis treats it so.

[12] The author acknowledges the wide range for the reported confidence intervals, which was a by-product of the sample size and represents a broader issue of research (Loewen & Hui, 2021). Based on the information required to calculate confidence intervals (i.e., *M*, *SD*, *n*), the author estimated that a sample size of approximately 300 would have been necessary to obtain tighter CIs, which would have simply not been feasible in research of this kind. However, in the face of a small sample size, following the guidance of Loewen & Hui (2021), a reasonably large amount of data was collected from each participant (63 items × 3 tests = 189 data points), resulting in 6615 data points in total (63 items × 3 tests × 35 participants), divided into data for the encountered (42 × 3 tests × 35 participants = 4410 data points) and unencountered cognates (42 × 3 tests × 35 participants = 2205 data points).

[13] However, the author acknowledges that this type of pedagogical approach (i.e., the use of historical linguistics) may be more suitable for specific types of learners—namely, those who are more analytic or have better phonemic coding capacities.

[14] The term applied historical linguistics has been used in different ways in the literature. Crystal (2016) used this term when describing efforts to teach Shakespearean pronunciation to stage actors. Campbell (2013) used it to refer to linguistic paleontology.

[15] The Second Germanic Sound Shift differentiates German not only from English but also from other Germanic languages. German-speaking L2 learners of Scandinavian languages, such as Norwegian, may benefit from the declarative knowledge that "t" became "ts" (orthographically "z") in German, accounting for differences such as Norwegian "tall [number]," "tid [time]," and "å betale [to pay]," and German "Zahl [number]," "Zeit [time]," and "bezahlen [to pay]."

REFERENCES

American Council on the Teaching of Foreign Languages. (2012). *ACTFL position statements*. American Council on the Teaching of Foreign Languages. http://nadsfl.org/wp-content/uploads/2017/04/ACTFL_Position_Statements.pdf

Anderson, R. C. (2004). Role of the reader's schema in comprehension, learning and memory. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (pp. 594–606). Erlbaum.

Arteaga, D. L., & Herschensohn, J. (1995). Using diachronic linguistics in the language classroom. *Modern Language Journal*, *79*, 212–222. https://doi.org/10.1111/j.1540-4781.1995.tb05432.x

Atkinson, R. C., & Raugh, M. R. (1975). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Journal of Experimental Psychology: Human Learning and Memory*, *1*, 126–133. https://psycnet.apa.org/doi/10.1037/0278-7393.1.2.126

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 89–195). Academic Press. https://doi.org/10.1016/S0079-7421(08)60422-3

Baddeley, A. (1990). *Human memory: Theory and practice*. Allyn & Bacon.

Barbour, S., & Stevenson, P. (1990). *Variation in German: A critical approach to German sociolinguistics*. Cambridge University Press.

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, *52*, 323–363. https://doi.org/10.1111/0023-8333.00186

Bolger, P., & Zapata, G. (2011). Semantic categories and context in L2 vocabulary learning. *Language Learning*, *61*, 614–646. https://doi.org/10.1111/j.1467-9922.2010.00624.x

Bower, G. H., & Clark, M. C. (1969). Narrative stories as mediators for serial learning. *Psychonomic Science*, *14*, 181–182. https://doi.org/10.3758/BF03332778

Campbell, L. (2013). *Historical linguistics* (3rd ed.). Edinburgh University Press.

Carpenter, S. K., Sachs, R. E., Martin, B., Schmidt, K., & Looft, R. (2012). Learning new vocabulary in German: The effects of inferring word meanings, type of feedback, and time of test. *Psychonomic Bulletin & Review*, *19*, 81–86. https://doi.org/10.3758/s13423-011-0185-7

Craik, F. I. M. (2002). Levels of processing: Past, present … and future? *Memory*, *10*, 305–318. https://doi.org/10.1080/09658210244000135

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Behavior*, *11*, 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*, 268–294. https://psycnet.apa.org/doi/10.1037/0096-3445.104.3.268

Craik, F. I. M., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *12*, 599–607. https://doi.org/10.1016/S0022-5371(73)80039-8

Crawford, R. L. (1988). Some aspects of the history of German and English for the classroom. *Die Unterrichtspraxis/Teaching German*, *21*, 204–207. https://www.jstor.org/stable/3530284

Crystal, D. (2016). Learning to be a linguist: The growth of linguistics in the 20th century. In A. G. Walker (Ed.), *Classics revisited: Wegbereiter der Linguistik neu gelesen* (pp. 195–212). Peter Lang.

De Wilde, V., Brysbaert, M., & Eyckmans, J. (2019). Learning English through out-of-school exposure: Which levels of language proficiency are attained and which types of input are important? *Bilingualism: Language and Cognition*, *23*, 171–185. https://doi.org/10.1017/S1366728918001062

Elgort, I. (2011). Deliberate learning and vocabulary acquisition in a second language. *Language Learning*, *61*, 367–413. https://doi.org/10.1111/j.1467-9922.2010.00613.x

Ellis, R. (2008). Explicit form-focused instruction and second language acquisition. In B. Spolsky & F. M. Hult (Eds.), *The handbook of educational linguistics* (pp. 437–455). Blackwell.

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3–25). Multilingual Matters. https://doi.org/10.21832/9781847691767-003

Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, *42*, 499–523. https://doi.org/10.1017/S0272263119000494

Gass, S. (1999). Discussion: Incidental vocabulary learning. *Studies in Second Language Acquisition*, *21*, 319–333. https://doi.org/10.1017/S0272263199002090

Gerrig, R., & Zimbardo, P. (2010). *Psychology and life* (19th ed.). Pearson.

Goo, J., Granena, G., Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443–482). John Benjamins.

Hall, C. J. (2002). The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *IRAL*, *40*, 69–87. https://doi.org/10.1515/iral.2002.008

Horsford, H. (1987). Common sense and language history. *Die Unterrichtspraxis/Teaching German*, *20*, 278–287.

Hu, H. M. (2013). The effects of word frequency and contextual types on vocabulary acquisition from extensive reading: A case study. *Journal of Language Teaching and Research*, *4*, 487–495. https://doi.org/10.4304/jltr.4.3.487-495

Hu, M., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *23*, 403–430.

Hulstijn, J. H. (1992). Retention of inferred and given word meanings: Experiments in incidental learning. In P. J. L. Arnaud & H. Benoint (Eds.), *Vocabulary and applied linguistics* (pp. 113–125). Macmillan. https://doi.org/10.1007/978-1-349-12396-4_11

Hulstijn, J. H. (2001). Intentional and incidental second language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258–286). Cambridge University Press.

Kang, E. Y., Sok, S., & Han, Z. H. (2019). Thirty-five years of ISLA on form-focused instruction: A meta-analysis. *Language Teaching Research*, *23*, 428–453.

Kirsch, C. (2012). Using storytelling to teach vocabulary in language lessons: Does it work? *The Language Learning Journal*, *44*, 33–51. https://doi.org/10.1080/09571736.2012.733404

Krashen, S. D. (2004). *The power of reading: Insights from the research*. Heinemann.

Krashen, S. D. (2011). *Free voluntary reading*. Heinemann.

Kroonen, G. (2013). *Etymological dictionary of Proto-Germanic*. Brill.

Larson–Hall, J. (2017). Moving beyond the bar plot and the line graph to create informative and attractive graphics. *Modern Language Journal*, *101*, 244–270. https://doi.org/10.1111/modl.12386

Lass, R. (1994). *Old English: A historical linguistic companion*. Cambridge University Press.

Laufer, B. (1988). The concept of 'synforms' (similar lexical forms) in L2 vocabulary acquisition. *Language and Education*, *2*, 113–132. https://doi.org/10.1080/09500788809541228

Laufer, B. (1997). The lexical plight in second language reading: Words you don't know, words you think you know and words you can't guess. In J. Coady

& T. Huchin (Eds.), *Second language vocabulary acquisition: A rational for pedagogy* (pp. 20–34). Cambridge University Press.

Laufer, B. (2005). Focus on form in second language vocabulary learning. In S. H. Foster-Cohen, M. García Mayo, & J. Cenoz (Eds.), *EUROSLA yearbook* (Vol. *5*, pp. 223–250). John Benjamins. https://doi.org/10.1075/eurosla.5.11lau

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

Lightfoot, D. (2007). Language history for teaching and learning German. *Die Unterrichtspraxis/Teaching German*, *40*, 34–45. https://doi.org/10.1111/j.1756-1221.2007.tb00028.x

Loewen, S., & Hui, B. (2021). Small samples in instructed second language acquisition research. *Modern Language Journal*, *105*, 187–193. https://doi.org/10.1111/modl.12700

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory and Cognition*, *29*, 10–17. https://doi.org/10.3758/BF03195736

Nagy, W. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. In R. Wagner, A. Muse, & K. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 52–77). The Guilford Press.

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

Nation, P. (2015). Principles guiding vocabulary learning through extensive reading. *Reading in a Foreign Language*, *27*, 136–145. https://doi.org/10.1125/66705

Norris, J. M., & Ortega, L. (2000). Effectiveness of L2 instruction: A research synthesis and quantitative meta-analysis. *Language Learning*, *50*, 417–528. https://doi.org/10.1111/0023-8333.00136

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*, 551–577. https://doi.org/10.1017/S0272263117000407

Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, *35*, 655–687. https://doi.org/10.1017/S0272263113000399

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912. https://doi.org/10.1111/lang.12079

Prince, P. (2012). Towards an instructional programme for L2 vocabulary: Can a story help? *Language Learning & Technology*, *16*, 103–120. https://doi.org/10.1125/44302

Ranalli, J. (2008). Learning English with *The Sims*: Exploiting authentic computer simulation games for L2 learning. *Computer Assisted Language Learning*, *21*, 441–455. https://doi.org/10.1080/09588220802447859

Rodgers, M. P. H., & Webb, S. (2020). Incidental vocabulary learning through viewing television. *International Journal of Applied Linguistics*, *171*, 191–220. https://doi.org/10.1075/itl.18034.rod

Roehr, K. (2018). *Metalinguistic awareness and second language acquisition*. Routledge.

Rott, S. (1999). The effect of exposure frequency on intermediate language learners' incidental vocabulary acquisition and retention through reading. *Studies in Second Language Acquisition*, *21*, 589–619. https://doi.org/10.1017/S0272263199004039

Salmons, J. (2012). *A history of German: What the past reveals about today's language*. Oxford University Press.

Schmitt, N. (2008). Instructed second language vocabulary learning. *Language Teaching Research*, *12*, 329–363. https://doi.org/10.1177/2F1362168808089921

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

Smith, S. (1968). Historical linguistics and the teaching of German. *The German Quarterly*, *41*, 231–238. https://doi.org/10.2307/403255

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning*, *60*, 263–308. https://doi.org/10.1111/j.1467-9922.2010.00562.x

Stedje, A. (2001). *Deutsche Sprache gestern und heute: Einführung in Sprachgeschichte und Sprachkunde* [German language yesterday and today: Introduction to language history and linguistics]. Wilhelm Fink Verlag.

Sundqvist, P. (2019). Commercial-off-the-shelf games in the digital wild and L2 learner vocabulary. *Language Learning & Technology*, *23*, 87–113. https://doi.org/10.1125/44674

Thornbury, S. (2016). Communicative language teaching in theory and practice. In G. Hall (Ed.) *The Routledge handbook of English language teaching* (pp. 242–255). Taylor & Francis.

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*, 46–65. https://doi.org/10.1093/applin/aml048

Webb, S. (2020). Incidental vocabulary learning. In S. Webb (Ed.), *The Routledge handbook of vocabulary studies* (pp. 225–239). Routledge.

Webb, S., & Chang, A., C.–S. (2015a). Second language vocabulary learning through extensive reading: How does frequency and distribution of occurrence affect learning? *Language Teaching Research*, *19*, 667–686. https://doi.org/10.1177/2F1362168814559800

Webb, S., & Chang, A. C.–S. (2015b). How does prior word knowledge affect vocabulary learning progress in an extensive reading program? *Studies in Second Language Acquisition*, *37*, 651–675. https://doi.org/10.1017/S0272263114000606

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Wolff, R. A. (1993). The history of the language as an instructional aid. *Die Unterrichtspraxis/Teaching German*, *26*, 27–35. https://doi.org/10.2307/3531786

Worthen, J. B., & Hunt, R. R. (2011). *Mnemonology: Mnemonics for the 21st century*. Psychology Press.

Yamamoto, Y. (2014). Multidimensional vocabulary acquisition through deliberate vocabulary list learning. *System*, *42*, 232–243. https://doi.org/10.1016/j.system.2013.12.005

Yates, F. A. (1966). *The art of memory*. Routledge & Kegan Paul.

APPENDIX A

Summary of Intentional Training Schedule[a]

Session	Content	Description
Session 1	Historical Linguistics	Explanation of the traditional narrative of 'English' being brought to Britain by three Germanic tribes (Angles, Saxons, Jutes) Instruction on Ingvaeonic Palatalization
Session 2	Sound Change	Review Instruction on Second Germanic Sound Shift
Session 3	Sound Change	Review on Second Germanic Sound Shift
Session 4	Semantic Change	Instruction on semantic changes
Session 5	Review	Review
Session 6	Review	Review

[a]A comprehensive breakdown of the schedule, along with example activities, is provided in the Online Supporting Information for this article.

APPENDIX B

Summary of Incidental Training Schedule

Session	Content	Description
Session 1	Communicative Activity	Two-way interaction task containing cognates and definitions. Learners communicated in German to find appropriate matches.
Session 2	Reading	Read short German text (250 words) containing some of the target words (+ distractors). Learners answered follow-up comprehension questions.
Session 3	Roleplay	Learners created a short roleplay based on a word list (target words + distractors) containing L2 definitions. Dictionary use was permitted.
Session 4	Heads-Up	Learners played a spinoff of the game "Heads up" (https://www.warnerbros.com/games-and-apps/heads). One learner held up a list of cognates and distractors and their partners had to provide them clues in the target language so they could guess it correctly.
Session 5	Speed Dating	Learners were given three words (target cognates + distractors) and they had a two-minute conversation (with ten different people) containing the words (e.g., target word = "Tier," response = "Was ist dein Lieblingstier [what's your favorite animal]?").
Session 6	Reading: Cloze Test	Short (250 word) reading containing some target cognates (+ distractors), followed by comprehension questions and a cloze test.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary material