Studies in Second Language Acquisition (2026), **48**, 808–832
doi:10.1017/S0272263126101831

RESEARCH ARTICLE

Learning cognates, false cognates, and noncognates contextually through reading: Do contextual support and cross-linguistic awareness matter?

Breno Silva[1], Małgorzata Foryś-Nogala[2], Katarzyna Kutyłowska[1], Jakub Szewczyk[3] and Agnieszka Otwinowska[1]

[1]Faculty of Modern Languages, Institute of English Studies, University of Warsaw, Warsaw, Poland; [2]School of Human Sciences, VIZJA University, Warsaw, Poland and [3]Institute of Psychology, Jagiellonian University in Kraków, Kraków Poland
Corresponding author: Agnieszka Otwinowska-Kasztelanic; Email: a.otwinowska@uw.edu.pl

(Received 07 November 2024 UTC; Revised 05 March 2026 UTC; Accepted 26 March 2026 UTC)

Abstract

Lexical acquisition often occurs through reading, yet little research has compared how cognates, false cognates, and noncognates are learned contextually through reading. To this end, we conducted a week-long pretest-training-posttest study with 54 adult Polish learners of English. During three sessions on consecutive days, participants read 45 short stories (170–317 words) containing 90 target keywords (30 cognates, 30 noncognates, 30 false cognates). We examined how word type, context informativeness, and number of keywords occurrences affected keyword learning. Additionally, participants were randomly assigned to a control or awareness-raising group, the latter trained to notice cross-linguistic similarities. Mixed-effects models showed learning of all word types at immediate posttest, with cognates being learned the most and false cognates the least. Higher context informativeness and more keyword occurrences improved learning for all keywords, but false cognates benefited the most from contextual clues. However, raising awareness of cross-linguistic similarity did not enhance contextual learning.

Keywords: vocabulary acquisition; orthographic similarity; cognates and false cognates; contextual learning; reading longer texts; awareness of cross-linguistic similarity

When reading in a foreign language (L2), learners encounter novel words repeatedly, which familiarizes them with the word forms and leads to learning. Learning from reading is often referred to as *incidental* or *contextual word learning* (henceforth: CWL; Elgort et al., 2018)[1]

[1]*Incidental* learning, often contrasted with *intentional* (i.e., deliberate) learning, may be defined as learning without the intention to commit words to memory, so participants are not informed of the aim of the experiment or the existence of a posttest (Laufer & Hulstijn, 2001). However, it may be impossible to ascertain

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

and may depend on the amount of information provided by the contexts where the words appear. The form of the words also matters, because novel words differ in how similar they are to the words in learners' native language (L1) and can be divided into three general word types: cognates, false cognates, and noncognates. Cognates bear orthographic/phonological and semantic similarities to words in learners' L1 (e.g., Polish *finalista* and English *finalist*), whereas false cognates (false friends, interlingual homographs; Otwinowska et al., 2020) are formally similar to but semantically different from their L1 counterparts (e.g., Polish *rumor*, meaning "a lot of noise," and English *rumor*, meaning "gossip").

Such cross-linguistic similarities, especially the orthographic ones, may impact L2-vocabulary learning from reading, thus making the acquisition[2] of cognates and false cognates an important focus of research. Even in unrelated languages such as Polish and English, the focus of this paper, the number of cognates exceeds 3,500 words due to borrowing (Otwinowska, 2015). Many of these cognates are vital for communication (e.g., Eng./Pol = *information/informacja*) and widely used. Thus, since cross-linguistic similarity is an important and common phenomenon, research on how it influences word learning across two languages can generalize to other language pairs. However, there is scarce research on cognate and false cognate learning via CWL through reading. Here, we present experimental research which addresses this research gap by investigating the CWL of Polish-English cognates, false cognates, and noncognates through the reading of strictly controlled short stories.

Intentional learning of cognates and false cognates

Existing evidence suggests that cognates are better known than noncognates and false cognates (Otwinowska & Szewczyk, 2019; Puimège & Peters, 2019; Silva & Otwinowska, 2019). One reason for this cognate advantage may be that the familiar L2 form and the form-meaning link for cognates are acquired faster than for other words (Marecka et al., 2021). As specified by the Ontogenesis Model (OM; Bordag et al., 2022) and the Parasitic Model (PM; Ecke, 2015) of L2 word acquisition, learners establish mental connections between L2 forms and L1 meanings, and links between L1 and L2 word forms, which evolve over time. The OM (Bordag et al., 2022) conceptualizes L2 word learning as a gradual refinement of representations—from initially "fuzzy" forms to more stable, native-like mappings within the L2. The PM (Ecke, 2015) posits that new L2 words "attach to" or "parasitize" pre-existing L1 formal and conceptual structures whenever L1-L2 similarity is detected. Thanks to parasitizing on L1 forms, the form-meaning mappings for L2 cognates are initially less "fuzzy" than for noncognates, so cognates are learned faster (Ecke & Hall, 2022).

However, intentional learning results for cognates are mixed. Cognates indeed have a learning advantage over noncognates when tested in isolation: They are easier to recall (Elias & Degani, 2022; Hirosh & Degani, 2021) and are less likely to be forgotten than noncognates (De Groot & Keijzer, 2000; Rogers et al., 2015). They are also easier to learn from supportive sentence contexts than other words (Mulder et al., 2019). Still, some classroom counterevidence exists. Rogers et al. (2015) found higher learning for noncognates than for cognates, and Otwinowska et al. (2020) found that cognates were

whether intentionality to learn novel words occurs during the reading process, or even post-experiment. Thus, following Elgort et al. (2018), we prefer the term CWL.

[2]In this paper we use "learning" and "acquisition" interchangeably.

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

acquired as much as noncognates and false cognates. Thus, it remains unclear whether a cognate learning advantage over other word types exists.

Studies exploring false cognates have also yielded mixed results. On the one hand, false cognates are translated less accurately and are less known than other word types (Janke & Kolokonte, 2015; Otwinowska & Szewczyk, 2019), which suggests that they may be harder to learn than cognates and noncognates. As suggested by De Groot (2011), learners initially link the L2 form to a wrong meaning, thus hampering the learning of false cognates. On the other hand, studies directly focusing on the learning of false cognates have not supported this hypothesis. Otwinowska et al. (2020) found no disadvantage of learning false cognates over cognates and noncognates in the classroom via intentional-learning vocabulary tasks. Similarly, Elias and Degani (2021) found no overall difference in learning between false cognates and noncognates.

By contrast, in a paired-associate learning study by Marecka et al. (2021), false cognates were learned slower than cognates but faster than noncognates. This was attributed to the faster learning of false-cognate forms, shared with their L1 orthographic neighbors. In line with the findings by Marecka et al. (2021), Hirosh and Degani's (2021) found a learning advantage of false cognates over noncognates when the German (L3) keywords were taught with English (L2) translations, but not with Hebrew (L1) translations. English is formally more similar to German than Hebrew, leading the authors to speculate, in agreement with Marecka et al. (2021), that form overlap facilitated learning more than meaning competition hindered performance, resulting in a false-cognate learning advantage over noncognates. As a result, even though Janke and Kolokonte (2015) and Otwinowska and Szewczyk (2019) found that noncognates are better known than false cognates, intentional-learning studies were unable to find the source of this false cognate disadvantage.

It is unclear whether cognates and false cognates are acquired less or more than noncognates. One possibility is that the cognate advantage and false cognate disadvantage may depend on whether words are learned intentionally or contextually. Contextual learning accounts for a significant part—if not most—of L2 vocabulary learning (Nation & Webb, 2011; Uchihara et al., 2019; Vidal, 2011) and will be discussed below.

Contextual learning of noncognates, cognates, and false cognates

Contextual word learning (CWL) is regarded as an essential source of L2 lexical learning (see Uchihara et al., 2019, for a meta-analysis). However, few studies have compared the CWL of cognates, noncognates, and false cognates. Vidal (2003) exposed her participants to English academic vocabulary through listening to lectures and measured the learning of three word types: cognates, noncognates, and "deceptively transparent words" (comprising three distinct word subtypes, false cognates being one of them). She found that cognates were learned the best, and noncognates the least, and it is unclear why deceptively transparent words registered more lexical gains than noncognates. Fievez et al. (2020), Peters and Webb (2018), and Peters (2019) compared the learning of cognates and noncognates through audiovisual input (audio with imagery) with or without onscreen text, finding that cognates were learned better than noncognates. Still, these studies investigated multimodal input, and it is uncertain whether results would be similar with reading-only input.

Finally, we are aware of only one study that investigated the impact of cognate status on CWL through reading. Vidal (2011) exposed learners to three academic texts and concluded that learning was facilitated by cognates but hindered by deceptively

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

transparent words. Similarly to Vidal (2003), false cognates were only one subtype of the deceptively transparent words category, making it impossible to statistically tease out the unique contributions of false cognates to learning. Clearly, given the conflicting results and the paucity of data, more research exploring factors that affect the learning of cognates, noncognates, and false cognates through reading is needed.

The influence of contextual support on word learning

CWL may be influenced by learner-related factors, including learners' L2 proficiency (Chen et al., 2017), working memory capacity (Elgort et al., 2018; Malone, 2018), and awareness of cross-linguistic similarity (Otwinowska, 2015). CWL also depends on many word-related factors, including cognateness, part of speech (Godfroid et al., 2017), polysemy, length, concreteness, and frequency in the L1 and L2 (Brysbaert & New, 2009). The two important word-related factors we will focus on here are the number of word occurrences and context informativeness.

Word Occurrences in CWL

Typically, more encounters with novel words in context result in more learning (Malone, 2018; Uchihara et al., 2019; Vidal, 2011). Some researchers suggest that around 10 encounters with a novel word are needed to learn its meaning (Chen & Truscott, 2010; Godfroid et al., 2017), although this may depend on the type of word knowledge being measured (Webb, 2007), word type, and the amount of semantic information provided in the contexts words appear in. We are unaware of any study exploring whether the number of novel-word occurrences in context impacts the learning of cognates, noncognates, and false cognates differently. Vidal (2011) explored cognateness and word repetition in CWL but did not model how these two aspects might interact.

Contextual Informativeness in CWL

Contexts are considered informative when they contain a high number of cues as to the meaning of novel words (Elgort et al., 2015), thus increasing the likelihood of accurate lexical inferencing and learning (Mulder et al., 2019). More informative contexts often result in faster L2 learning and higher lexical gains than less informative contexts (Chen et al., 2017; Elgort et al., 2015; Mulder et al., 2019; Webb, 2008). Mulder et al. (2019) investigated the learning of L2-English words embedded in sentences with different degrees of contextual informativeness. Their participants more accurately translated words placed in highly informative contexts. Similarly, Webb (2008) explored CWL by measuring the learning of L2 words in short contexts (one or two sentences) while participants performed a reading-comprehension task. Consistently with Mulder et al. (2019), lexical learning was higher among participants exposed to more informative contexts. By contrast, no effect of contextual informativeness on lexical processing and learning was found in an eye-tracking study with words embedded in sentential contexts (Yi et al., 2022). However, it is unclear what would happen when keywords are embedded in longer texts. Also, and critically, no study has investigated whether contextual informativeness affects cognates, noncognates, and false cognates differently.

Awareness of L1-L2 similarity in the learning of cognates and false cognates

The learning of cognates and false cognates in CWL might also be positively influenced by learners' heightened awareness of cross-linguistic similarity (Otwinowska, 2015;

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

White & Horst, 2012). Drawing on Schmidt's (1990) Noticing Hypothesis and its extension to vocabulary by Laufer and Hulstijn (2001), Otwinowska et al. (2020) posited that awareness of cross-linguistic similarity operates at the level of noticing and understanding. They operationalized "noticing" as the ability to identify the L2-L1 cognates in a text, which requires attentional focus, and "understanding" as explicit knowledge or verbalization of the language rules underlying L1-L2 similarity.

Raising learners' awareness at the level of understanding often results in noticing cognates better, although the existing results are mostly qualitative (e.g., White & Horst, 2012), and better noticing may enhance learning (Schmidt, 1990). Therefore, it is possible that if language learners are trained to consciously attempt to identify cognates and false cognates in input, learning of these word types may be affected. Otwinowska et al. (2020) were the first to empirically test whether awareness of cross-linguistic similarities may influence L2 learners' vocabulary learning. However, they found no difference between the high- and low-awareness groups in the intentional learning of noncognates, cognates, and false cognates in class. They hypothesized that if awareness raising is to confer any advantage to cognate learning, it might occur in CWL, which involves less intense attentional focus on the keywords than classroom activities.

Limitations in previous research

We have drawn attention to several research gaps. First, few studies on cognates and false cognates have investigated their contextual learning. Most have used audiovisual input and excluded false cognates from their design (e.g., Peters, 2019; Peters & Webb, 2018). Only two studies have investigated CWL of these word types (Vidal, 2003, 2011), and only Vidal (2011) explored CWL through reading, but measured false-cognate learning only indirectly. Second, it is unclear if the number of novel word occurrences in context impacts the learning of cognates, noncognates, and false cognates differently. Also, it is unclear how context informativeness influences learning of these word types when reading texts longer than a sentence or two (e.g., Mulder et al., 2019; Webb, 2008).

Finally, it appears that many studies did not control for cognate guessing, including studies using meaning-recall (L2-into-L1 translation) tests, as we do in the current study. Meaning-recall tests are appropriate and accurate measures of receptive vocabulary knowledge (Nation & Webb, 2011). However, they are susceptible to guessing and score inflation, especially when testing the knowledge of cognates (Otwinowska & Szewczyk, 2019; Silva & Otwinowska, 2019). The current study addresses all the above-mentioned gaps.

The current study

Our study used tailor-made short stories where 30 cognate, 30 false-cognate, and 30 noncognate keywords were embedded in a controlled manner. To measure the contextual word learning (CWL) of the 90 keywords, we controlled for the number of keyword occurrences and the amount of contextual informativeness for each keyword. In agreement with Vidal (2011), we expected that fewer repetitions may be necessary to learn cognates than noncognates and false cognates. We also predicted that false cognates may rely on contextual informativeness the most, while cognates require little confirmation from the context to be acquired.

Participants read the stories for comprehension in three sessions that were run on consecutive days. Prior to each reading session, we manipulated participants' awareness

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

of cross-linguistic similarity. Using a pretest-posttest design, we measured gains in receptive word knowledge via a meaning-recall (L2-to-L1) test. To help control for cognate guessing in tests, we added a confidence scale for each keyword translation (see Methods). We asked the following research questions:

RQ1: Are there any differences in the learning of cognates, false cognates, and noncognates contextually through reading?

RQ2: Does contextual support (i.e., contextual informativeness and number of keyword occurrences) promote the CWL of noncognates and false cognates more than of cognates?

RQ3: Does awareness-raising impact the CWL of cognates, false cognates, and noncognates differently?

Procedure overview

The experimental procedures were approved by the Rector's Committee for the Ethics of Research Involving Human Participants at the University of Warsaw, Poland. All data and materials can be found in Appendices 1–14, openly available at https://osf.io/vfzwb/overview.

We provide the procedure overview first so that the consecutive steps concerning study participants and instruments are clearer. The procedure lasted 5 days, Friday to Thursday (see Figure 1), and each session lasted around 90 minutes.

After the online screening (background questionnaire, proficiency measures, pre-procedure Vocabulary Learning Strategies questionnaire), eligible participants (see *Participants*) were invited to take part in the project. Participants were ensured that the purpose of the study was to investigate the process of reading comprehension in English and were never told about the existence of a posttest. During Day 1 (Friday), participants signed consent forms, took the vocabulary pretest on 230 words, and background measures (Cambridge Placement Test for Schools and inhibition/aptitude tasks: Simon and written nonword repetition; NWR). Over the weekend, based on the pretest, we excluded those participants whose knowledge of our 90 keywords (30 cognates, 30 false cognates, and 30 noncognates) was too high and knowledge of 80 story words (*not* the keywords, but easy nouns taken from the stories) was too low (see *Participants* and *Materials*).

On days 2, 3, and 4 (Monday–Wednesday), the remaining participants were randomly assigned to either the experimental or control condition. Each day involved a) Experimental Manipulation: raising awareness of cross-linguistic similarity for the experimental group, dummy training for the control group, and reading strategy training for both groups to mask the research purpose (see *Experimental Manipulation*); b) Reading Sessions 1, 2, and 3 (15 stories each day; see *Main study: Reading task*). At the end of each Reading Session, participants completed an additional grammatical or aptitude task, which served as additional distractors. On Day 5, participants took the vocabulary posttest, the awareness measures (Cognate Noticing Task, post-procedure Vocabulary Learning Strategies questionnaire), the Digit Span test, and had a Debriefing Session.

Participants

Participants were members of the general public, speakers of L1-Polish, recruited through a job portal in Poland. The online pre-screening included an online background questionnaire (L1, other languages, Vocabulary Learning Strategies

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

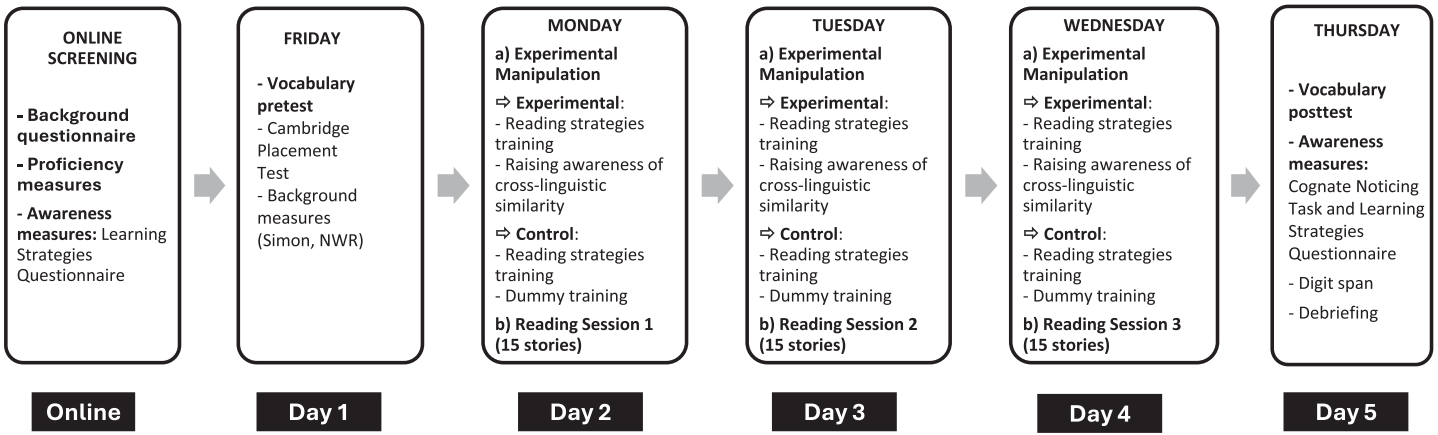


Figure 1. The main stages of the experimental procedure.

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

Questionnaire) and online proficiency tasks: the English LexTALE (Lemhöfer & Broersma, 2012) and the Cambridge Placement Test for general English (*Cambridge English Proficiency Test*, n.d.). After the pre-screening, 135 out of 227 L1-Polish participants were invited for the first session (see the *Procedure* section). After the admission to the study, participants were further screened regarding their knowledge of the 90 keywords (see *Materials; Keywords*) and 80 story words in our short stories (i.e., easy nouns selected from the stories, K1-K2 levels in Vocabprofile; Cobb, 2026). We required all participants to know at least 70% of the story words confidently (*Mdn* = 92.5%). The threshold values for participants selected for the study were a joint score on the LexTALE (minimum 50%, maximum 88%, *M* = 68.1; intermediate level) and Cambridge (minimum 10, maximum 22, *M* = 18.1; intermediate level). We aimed for intermediate-level participants but accepted those with slightly unbalanced results (e.g., lower on LexTALE and higher on Cambridge, or vice versa). We also excluded participants who knew more than 75% of all the keywords, as they would not show detectable effects of learning. We did not exclude anyone based on their knowledge of languages other than L1-Polish and L2-English. All 135 participants received financial compensation for this part of the experiment.

Consequently, 72 participants (keywords known *Mdn* = 45%, *range*: 15–72%) were invited to take part in the main experiment (see Appendix 1 for all participants' data). One participant was excluded due to a technical error, two withdrew from further participation, and fifteen were removed due to unreliable confidence ratings (more than 50% of incorrect, form-based translations of false cognates were done with full confidence). The final sample consisted of 54 speakers of L1-Polish (M_{age} = 22.7, *SD* = 4.8, *range*: 18–39) randomly assigned to either experimental or control group (27 participants each group). Overall, the participants' proficiency in L2-English was B2+ CEFR (Council of Europe, 2001). As shown in Table 1, the experimental and control groups did not differ significantly on English proficiency, knowledge of story words, reading comprehension skills (see *Instruments*), working memory capacity (Digit Span sub-test from the Polish version of the Wechsler Adult Intelligence Scale; WAIS-R; Brzeziński et al., 2004), language learning aptitude (written nonword repetition task), and cognitive inhibition (Simon task).

Materials

Keywords

We used the same keywords as those used in the classroom study by Otwinowska et al. (2020). They were 90 English nouns: 30 noncognates, 30 Polish-English cognates, and

Table 1. Descriptive Statistics for Background Participant-Related Measures Used in the Study

Measure	Experimental Group (*n* = 27) *M* (*SD*)	Control Group (*n* = 27) *M* (*SD*)	Eta squared[1]
LexTALE (max. 100)	68.96 (9.58)	70.50 (9.92)	.006
Cambridge Placement Test (max. 25)	19.22 (2.69)	19.07 (2.60)	.001
Reading Comprehension Task (max. 8)	5.85 (1.75)	5.37 (1.74)	.019
Digit Span (max. 28)	14.96 (4.29)	14.22 (2.71)	.011
Knowledge of story words (%)	.92 (0.05)	.92 (0.06)	0
Simon task	.09 (0.03)	.08 (0.02)	.015
Written nonword repetition task (max. 50)	43.93 (5.92)	43.04 (5.80)	.006

[1]Effect size thresholds for applied linguistics are between .06 to .15 (weak), .16 to .35 (moderate), and >0.36 (large) (Plonsky & Oswald, 2014).

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

Table 2. Types of Keywords Used in the Stories

Concreteness	Cognates (*n* = 30)	False cognates (*n* = 30)	Noncognates (*n* = 30)
More abstract (*n* = 45)	*Arrogance* (Pl. arogancja)	*Concurrence* (Pl. konkurencja; Eng. translation: competition)	*Achievement* (Pl. osiągnięcie)
More concrete (*n* = 45)	*Institution* (Pl. instytucja)	*Pension* (Pl. pensja; Eng. translation: wages)	*Sculpture* (Pl. rzeźba)

Note: Abbreviations: Eng. = English, Pl. = Polish.

Table 3. Descriptive Statistics for the Keywords Used in the Stories and Their Context Informativeness

Measure	Cognates *M (SD)*	False cognates *M (SD)*	Noncognates *M (SD)*	Eta squared[1]
Concreteness	3.00 (1.08)	3.12 (1.12)	3.12 (.85)	.003
Log_{10} L2 Frequency	2.20 (.45)	2.29 (0.65)	2.36 (.43)	.017
nLD (normalized Levenshtein Distance)	.25 (.12)	.28 (1.79)	_	.009
Word length	8.23 (2.24)	8.40 (1.52)	7.20 (2.31)	.065
Number of Keyword occurrences	6.20 (2.44)	6.07 (2.88)	5.73 (2.49)	.006
Contextual informativeness (GPT2)	−5.57 (1.66)	−6.54 (2.44)	−4.64 (1.99)	.128
Contextual informativeness (CP)	.35 (.17)	.33 (.16)	.35 (.16)	.003

[1]0.06 to 0.15 (weak), .16 to .35 (moderate), > 0.36 (large) (Plonsky & Oswald, 2014).

30 Polish-English false cognates chosen based on their orthographic and semantic overlap (Table 2; see Appendix 2 for all keywords and item-related variables). The matching procedure and rationale are described in the original publication.

In short, we selected English words not typically introduced before B1/B2 level so that they were not too easy for the participants. Additionally, the three word types were matched for frequency of occurrence in the L2 (SUBTLEX-US; Brysbaert & New, 2009); concreteness (Brysbaert et al., 2014), length, and, in the case of cognates and false cognates, L2-L1 orthographic similarity (normalized Levenshtein distance). The variables had very similar values across the three word types (Table 3) but were still included in the main regression model as covariates to reduce error. Additionally, the three word types were controlled for the number of occurrences across the input material (see *Number of Keyword Occurrences*) and the informativeness of their context (see *Contextual Informativeness*).

The input for contextual learning: 45 stories for reading sessions 1, 2, and 3

The 45 stories were composed as follows. We created 45 sets of keywords (one for each story) by randomly sampling from all 90 keywords so that each set contained 6 unique keyword types (see *Number of Word Occurrences* for distribution of keywords across stories). All keywords were treated as nouns (e.g., *sacrifice* could be a noun and a verb); inflections were allowed (e.g., plurals) but not derivations.

Stories were written by students and faculty members of the Institute of English Studies, University of Warsaw, Poland (one to four stories per author), which promoted thematic and stylistic variation. All drafts were revised and edited for coherence, accuracy, and naturalness, and proofread by a professional translator native speaker

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

of English. Still, given the need to embed often-infrequent keywords (e.g., *concurrence*), the stories were inevitably unnatural at some points (see Chen & Truscott, 2010, for a similar rationale). We made sure that all keywords had the same meaning across the stories, appeared in different sentence positions, and were separated from one another by at least one clause (preferably a sentence) to provide enough understandable context. Next, any slang or idiomatic expressions, or other items that could hamper text comprehension were changed into simpler language.

All stories were analyzed with Vocabprofile (Cobb, 2026) to ascertain that all running words—except for the keywords—were among the 2,000 most frequent word families in English. Importantly, the keyword tokens did not exceed 5% of the total word count (M = 4.7). Therefore, at least 95% of the story words were assumed to be easily understandable to participants, the minimum coverage necessary for learners to attain satisfactory levels of comprehension and perform lexical inferencing accurately (Hu & Nation, 2000). The 45 stories totaled 11,511 tokens, varying in length from 170 to 317 tokens (M = 256), with a Flesch Reading Ease mean of 74.94 (SD = 7.42), thus further indicating that the stories were fairly easy to read. An excerpt from a representative story is provided below (keywords in bold; see Appendix 4 for all stories):

> Last week, I attended a fascinating **lecture** on female characters in British books of the last few centuries. Professor Davenport talked about women's characters (**heroines**) in the works of Jane Austen, a very famous female writer (…)

Estimates of contextual support for the keywords

Number of keyword occurrences

Each type occurred 1–4 times per story, which amounted to 10–15 occurrences of the keywords (tokens) per story. For instance, the following set, showing 13 keyword tokens, comprised six keyword types (number of tokens in brackets): *intellect* (1); *manifestation* (2); *lecture* (2); *heroine* (4); *minority* (3); *discipline* (1). In total, 2–10 tokens of each keyword occurred across the 45 stories, and each keyword appeared in 1–6 different stories (see Appendix 3 for the keyword distribution across the stories). Due to a technical error, one keyword, *budget*, appeared in seven different stories. All three word types had similar mean number of occurrences and SDs across the stories (M = 5.84–6.23, SD = 2.49–2.86). A One-way ANOVA showed no difference between the word types: $F(2, 87) = .180$, $p = .836$, $\eta^2 = .004$.

Contextual informativeness

For each occurrence of each keyword in the stories, we obtained two estimates of how predictable the meaning of the keyword was given the surrounding context: Cloze probabilities (CPs; Taylor, 1957) and probabilities estimated by an AI tool (GPT-2).

Calculating cloze probability

CP provides a measure of contextual informativeness for each keyword at its position in a sentence and has been used extensively to assess predictability of words in context (e.g., Szewczyk & Federmeier, 2022). Higher CP scores indicate higher likelihood of successful contextual inferencing. To calculate CP for each keyword, we recruited

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

native speakers (NSs) of English from the UK and USA (via *Amazon Mechanical Turk* n.d.) with at least high-school-level education. They filled in gaps at the keyword positions with the first word that came to their mind. Each NS was presented with one gap at a time. They read the entire part of the story preceding the gap, which was followed by the rest of the sentence. If the gap occurred in sentence-final position, an extra sentence was provided.[2] After filling the gap, the NS saw the story fragment with the actual keyword, and the next part of the story until the next gap. Data were collected until each gap had been filled by at least 20 NS (see Szewczyk & Federmeier, 2022, for a similar procedure). The mean CP across all gaps was .35 (SD = .29), indicating moderately high contextual support. This means that, on average, the context provides enough clues about the word's meaning that 35% of NS continued the sentence with the keyword. Of note, in 31% of contexts, the gap was filled correctly by 50% of NS.

GPT-2 log probability

A potential limitation of CPs is that they poorly quantify the amounts of contextual support for words that are plausible but infrequent (Szewczyk & Federmeier, 2022). This is because infrequent words are unlikely to be chosen by native speakers when filling the gaps. In such cases, CPs could be zero, even though the context is still informative (i.e., some words are more compatible as continuations to the gap than others; see Appendix 5 for more details).

To overcome this limitation, following Szewczyk and Federmeier (2022), we used an alternative index of contextual support obtained from the AI tool GPT-2 (gpt2-xl; Radford et al., 2019; using Transformers library from HuggingFace; Wolf et al., 2020). GPT-2 is a neural network model trained on a very large corpus of texts. We employed GPT-2 to obtain (log-transformed) probability measures similar to cloze probability, but superior in that they better quantify contextual support for infrequent words: GPT-2 log-probability scores capture contextual semantic cues for the keywords even when CPs would be 0, and it does so for every word in the lexicon at any position in the sentence.

To assess the effectiveness of CP and GPT-2 log-probability, we created three models: One including both measures (correlated at r = .67), and two others, each with only one of the variables. We then compared these models using an ANOVA to determine whether either measure explained additional variance beyond the other. GPT-2 log-probability emerged as a better predictor of lexical learning: Removing GPT-2 from the full model worsened model fit (p < .05), whereas removing CP did not (p = .66). Therefore, we used GPT-2 scores as the index of contextual informativeness.

Instruments

Main study: Reading Task

Participants read our 45 stories (see *The Input for Contextual Learning*) in six randomized orders. Within each randomization order, a participant read 15 stories per Reading Session (days 1, 2, and 3; Monday–Wednesday; see Figure 1). Ultimately, each participant was exposed to all 45 stories in one of the six orders.

Each story was presented in full on a screen. The presentation was timed, as established during piloting (600 ms per word; e.g., 147 seconds for a story consisting of 247 words). If participants finished reading earlier, they could press the spacebar to move to comprehension questions (three multiple-choice questions per story,

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

introduced to ensure that participants read attentively). To avoid inadvertent exposure, keywords were avoided in the questions and answers.[3]

L2-L1 translation pretests and posttests with confidence ratings

To measure participants' gains in lexical knowledge, we used identical L2-L1 translation pretests and posttests. The test incorporated 230 items: 90 keywords (nouns; 30 cognates, 30 false cognates, and 30 noncognates, see *Keywords*), 30 control words, 80 story words, and 30 nonwords (generated from the keywords using the default settings in the Wuggy software; Keuleers & Brysbaert, 2010). The nonwords were introduced as fillers to discourage wild guessing. The 30 control words were matched to the keywords in L2 frequency, concreteness, and length. They did not appear in the stories and were used to control for the test-retest effects (i.e., any learning stemming from taking the same test twice). The 80-story words (easy nouns used in the stories, K1-K2 levels in Vocabprofile; Cobb, 2026) were included to ensure that the stories were understandable to readers as one of the key inclusion criteria in the study (see *Participants*), and to divert learners' attention away from the keywords in the pretest.

The test was computerized and the order of the 230 items was randomized for each participant. Participants first assessed whether each of the 230 forms presented on the screen was an existing English word. Next, participants wrote Polish translations of the words they considered real words. For each translation, they rated their confidence on a 7-point Likert scale (from 1 *I'm guessing* to 7 *I know for sure*), which was used to control for guessing (see Data analysis)

Experimental manipulation: Training tasks for awareness raising

During Days 2, 3, and 4 (Monday–Wednesday; see Figure 1), before each Reading Session, participants were told that they would receive a mini-lesson aimed at assisting them with text comprehension (on average 10 minutes). The computerized mini-lessons, self-paced, served to introduce the Experimental Manipulation and were similar in structure and length for the experimental and the control groups (see Appendix 6 for all activities). Each mini-lesson started the same for all participants and concerned reading-comprehension strategies (to participants' knowledge, the goal of the study). Next, the experimental group focused on cognates and false cognates. Experimental group participants were advised that paying attention to cognates might help them understand L2 texts and were warned about false cognates. They also completed several cognate and false cognate identification and translation tasks, and were given immediate feedback. The Experimental Manipulation (5–7 minutes per day) totaled around 20 minutes of intensive awareness-raising of cross-linguistic similarity. The control group did a dummy training of similar structure and length on compounds and morphologically complex words. None of the words used in the mini-lessons (experimental or control) were used as keywords in the stories. To measure the effectiveness of such training, we used the two awareness measures described below.

[3]In rare cases in which two keywords occurred in the same sentence, the gap was followed by all words of a given phrase (to avoid giving away the word occurring at the position of the next gap or presenting two gaps in one item).

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

Awareness measures

Vocabulary Learning Strategies Questionnaires

Before and after study participants filled in a questionnaire on Vocabulary Learning Strategies (see Appendix 1 for a list of strategies). One strategy was "Transfer," defined as "consciously seeking similarities between L1 and L2 words." Answers to the question regarding Transfer were used to assess whether the Experimental Manipulation (i.e., awareness raising of cognates and false cognates) changed participants' learning behavior.

Reading Comprehension and Cognate Noticing Tasks

The tasks were both based on one reading text of 375 words in English (see Appendix 7), and both were performed post-experiment. The text included 69 Polish-English cognates which had not been used in the stories, the main task.

First, participants read the text and responded to eight open-ended comprehension questions (scoring = 1 point each; Cronbach's alpha = .55, 95% CI [.347, .710]). This Reading Comprehension Task was done to ensure the groups did not differ in terms of reading comprehension skills, which was confirmed (p = .40; see Table 1).

Afterward, using the same text, participants performed the Cognate Noticing Task. They were asked to underline all words they considered similar to Polish. Here, we tested participants' ability to notice novel cognates in written texts and thus verify whether the awareness-raising manipulation had been effective (in addition to the Vocabulary Learning Strategies questionnaire described above). The reliability of the Cognate Noticing Task yielded Cronbach's alpha = .931, 95% *CI* [.902, .955].

Posttest and debriefing interviews

On Day 5 (Thursday), participants took the computerized vocabulary posttest (same as the pretest), the Cognate Noticing Task and the Vocabulary Learning Strategies questionnaire (to measure their awareness of cross-linguistic similarity post-hoc), and the Digit Span working memory test. Finally, following suggestion by De Vos et al. (2019), participants were debriefed via a semi-structured interview (see Appendix 8) to ascertain that the learning generated by the stories was not intentional. Thirty-one participants claimed to know the purpose of the experiment (vocabulary learning), despite our efforts to hide its true aim, and 33 participants looked up some words outside the experimental session (M = 1.94, SD = 1.22, range = 1–5 words out of 90 keywords). To avoid biasing results, the relevant data points were removed from the statistical analyses.

Data analysis

Correction for guessing

Asking participants to translate L2 cognates and false cognates may encourage guessing. Among false cognates, guesses are easy to detect because they lead to incorrect translations. A preliminary analysis confirmed the extent of the problem: 76% of all incorrect false cognate translations were guessed based on word form, i.e., participants assumed the word was a Polish-English cognate. Thus, we concluded that participants may have also guessed cognates based on form, inflating cognate scores. Still, cognate guessing is hard to detect because it results in correct answers. A workaround was

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

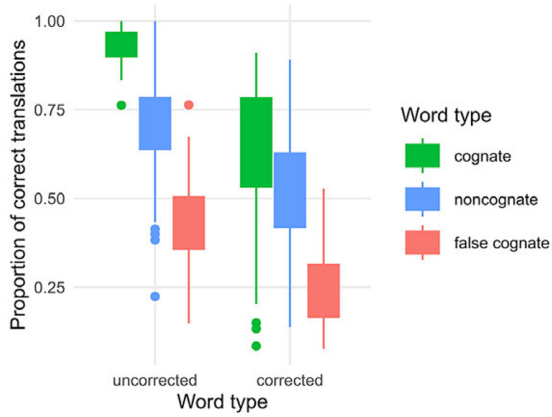


Figure 2. Proportion of correct translation before and after applying the correction for guessing. Data are aggregated across pretest and posttest.

suggested by Otwinowska and Szewczyk (2019) and Otwinowska et al. (2020), who hypothesized that guessed answers are made with lower confidence than answers based on true knowledge. Following these papers, we considered as correct only translations given with full confidence (i.e., 7 on a scale from 1 *I'm guessing* to 7 *I know for sure*), while the remaining translations were treated as incorrect.

This approach was effective in our data. Limiting correct responses to full-confidence translations reduced the proportion of guessed false-cognate translations —i.e., translated as cognates—from 76% to 27%. As expected, this correction disproportionately affected cognates (see Figure 2), whose translation accuracy declined twice as much as that of other word types. We therefore applied the correction in all subsequent analyses (although see Appendix 13 for alternative approaches).

Statistical modeling

We operationalized word knowledge as the ability to translate the L2 word into the L1 with full confidence (confidence = 7, *I know for sure*), similarly to Otwinowska et al. (2020). However, in Appendix 13, we conducted a more fine-grained analysis, separating learning into two subcomponents: accuracy-driven (words that went from incorrect to correct with full confidence) and confidence-driven (words that went from correct with low confidence to correct with full confidence). This allowed us to distinguish between gains in form-meaning mapping (i.e., our operationalization of learning) and increases in certainty about that mapping. The conclusions in Appendix 13 support the conclusions in this paper.

The dependent variable was binomial since the answers were coded either as correct or incorrect. All missing responses, errors, and misspellings were treated as incorrect. Because the scores for the three word types differed significantly at pretest (see *Results*), analyses of learning focused on posttest results and were limited to items unknown at pretest. This approach allowed us to isolate words that could be learned and test whether they had been acquired through reading the stories. All models included control covariates accounting for variance unrelated to the variables of interest. There

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

Table 4. Average Proportion of Word Types Known Before and After the Experiment

	Experimental group		Control group	
Word type	Pretest	Posttest	Pretest	Posttest
Cognate	.63	.66	.59	.67
False cognate	.20	.31	.19	.30
Noncognate	.48	.56	.45	.58
Control word	.30	.29	.27	.31

Note. These are raw means (after the correction for guessing), and not the predicted results after modeling.

were nine keyword-related and six participant-related covariates (see Appendix 12 for the descriptive statistics of the item-related variables; see Table 1 for the participant-related covariates). In the main text we report the variables of interest only. See Appendices 9–11 for full models.

The CWL data were analyzed using generalized linear mixed-effects models with a logistic link function. The models estimated the probability of correctly translating the word. We included random intercepts and slopes for all effects of interest, but to avoid convergence issues, we did not include random correlations. To limit the number of estimated parameters, we did not include random slopes for the control variables (as we are not interested in precise estimates of their effect sizes or significance). All predictors were centered. Categorical predictors were deviation-coded. All raw data and scripts reproducing the analyses can be found online at: https://osf.io/vfzwb/overview

Results

Descriptive statistics and pretest results

Table 4 presents the descriptive statistics illustrating the average proportion of each word type known by the participants from the experimental and control groups.

Before addressing the research questions, we first assessed baseline word knowledge at pretest. The analysis used Word Type as the main predictor, with control variables included. As shown in Figure 3, cognates were significantly better known than noncognates (Δlogodds = .89, 95% CI [.07, 1.7], $p < .05$). In contrast, false cognates and control words were known substantially less well than noncognates (false cognates: Δlogodds = −1.73, 95% CI [−2.52, −.94]; $p < .0001$; control words: Δlogodds = −1.21, 95% CI [−1.99, −.43]; $p < .01$). The full model is presented in Appendix 9.

Which word types are easier to learn contextually?

RQ1 asked whether there were differences in the contextual learning of cognates, false cognates, and noncognates. The main analysis included Word Type (noncognate, cognate, false cognate, control word), Group, and all control predictors. The results show CWL for all previously unknown word types (learning rates significantly above 0, $ps < .0001$; see Figure 4) Control words—the only type not included in the contextual learning phase—had a probability of learning slightly above zero ($prob_{learn}$ = .05; 95% CI [.03, .8]), but this was significantly smaller than that of noncognates ($prob_{learn}$ = .25; 95% CI [.17, .34]; $p < .0001$), possibly reflecting the testing effect. Compared to noncognates, cognates were learned better ($prob_{learn}$ = .39, 95% CI [.28, .52]; $p < .05$), while false cognates were learned significantly worse ($prob_{learn}$ = .13, 95% CI [.09, .19], $p < .01$). The full model is presented in Appendix 10.

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

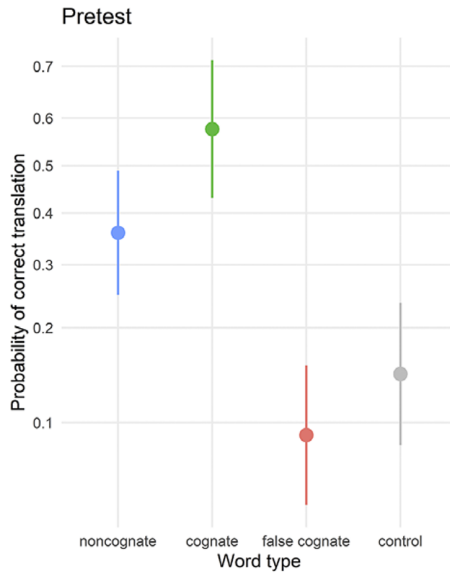


Figure 3. Modeled predictions for initial word knowledge. The bars indicate 95% confidence intervals.

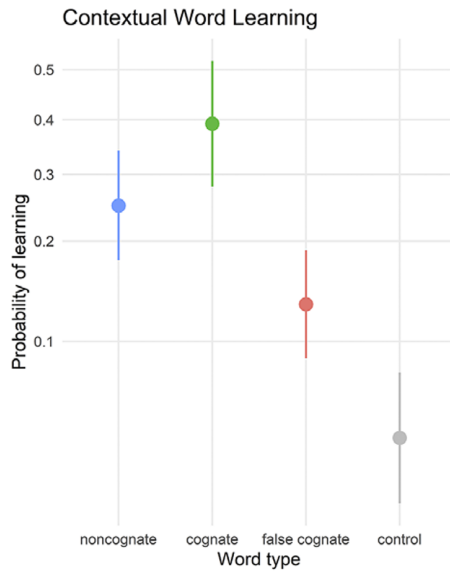


Figure 4. Modeled predictions for contextual learning. The bars indicate 95% confidence intervals.

In the analysis above, we focused on words that were unknown at pretest (only those could show learning), and we considered as learned only those that were correctly translated with full confidence in the posttest (our correction for guessing).

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

In Appendix 13 we checked if the learning effects were driven by increases in translation accuracy or in confidence (for words already translated correctly at pretest). These exploratory analyses confirmed our results. Briefly, they show that the pattern was primarily driven by differences in translation accuracy across word types, while increases in confidence were seen similarly across all word types and appeared to result partly from the testing effect.

Does contextual support promote the CWL of cognates, noncognates, and false cognates differently?

RQ2 asked whether the quality of contextual support across occurrences (i.e., mean context informativeness about the keyword) and its quantity (the number of times the target word appeared across stories) modulate CWL for each word type. For this and subsequent analyses, we fitted a single model that included Word Type, Group (Experimental vs. Control), Contextual Informativeness, and Number of Occurrences, plus the interactions of Word Type with the three other predictors. All control variables were included as main effects only. The results are shown in Figure 5; the full model is provided in Appendix 11.

The results show that mean context informativeness influences the likelihood of learning words of all types (B = .13; p = .045, 95% CI [.01, .25]), but false cognates are influenced by context informativeness the most (B = .31, p = .013, 95% CI [.07, .56]). Specifically, false cognates are harder to learn when the context offers weak semantic cues (Figure 5, left). The same model also shows that all word types benefit from repeated exposures to the target word (B = 1.77, p < .0001, 95% CI [.96, 2.58]; Figure 5, right), with no interaction with Word Type. The full model is presented in Appendix 11.

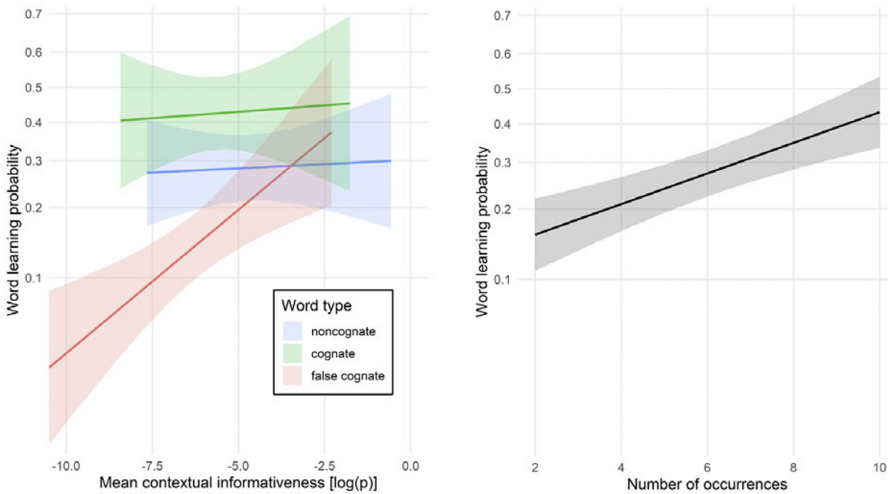


Figure 5. Analysis of word-learning predictors.
Note: Left: The effects of Mean Contextual Informativeness (higher value in the X-axis indicates the context is more informative). For examples of sentences and target words with different log(p) values, see Appendix 14. Right: The effect of the number of occurrences.

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

Table 5. Cognate Awareness Measures and Group Comparisons

Awareness measure	Experimental Group (n = 27) M (SD)	Control Group (n = 27) M (SD)
Transfer strategy pretest (max 7)[1]	3.80 (1.92)	3.97 (2.05)
Transfer strategy posttest (max 7)[1]	4.83 (1.76)	4.70 (1.89)
Awareness posttest (max. 69)[2]	34.7 (12.68)	29.7 (13.61)

[1]"Transfer" strategy obtained from the Vocabulary Learning Strategies Questionnaire.
[2]Cognate tokens underlined in the Cognate Noticing Task.

Does awareness-raising impact the CWL of cognates, false cognates, and noncognates?

Finally, RQ3 asked whether awareness-raising impacted the learning of cognates, false cognates, and noncognates differently. We addressed this question in three ways. Table 5 presents the descriptive statistics for the Experimental and the Control groups used in the analyses.

First, using the model from the previous subsection, we estimated the effect of Group (experimental vs. control), and its interaction with Word Type. We found no main effect of Group (B = .01, p = .97, 95% CI [−.40, .41]) and no interaction with Word Type (ps > .08; for details, see Appendix 11), yielding no support for the idea that awareness-raising training improves vocabulary acquisition (either overall, or for any specific word type).

Second, using the Vocabulary Learning Strategies questionnaire, we compared the self-reported use of the Transfer strategy, defined as "consciously seeking similarities between L1 and L2 words," before and after the experiment. Overall, the reported use of Transfer increased significantly from the pretest to the posttest (p < .001, 95% CI: [.46, 1.46]), but it did not interact with Group (experimental vs. control group; 95% CI: [−.85, 1.15]; p = .77).

Third, the Cognate Noticing Task showed that participants in the Experimental group underlined on average more cognate types than participants in the Control group (t(52) = 2.78, p = .007, 95% CI [.17, .97], Cohen's d = .76, representing a moderate effect size; Plonsky & Oswald, 2014).

Summing up, the experimental group was slightly better than the control group at noticing cognates when prompted, which indicates their increased awareness at the level of noticing. However, given the nonsignificant results for the Transfer strategy, the training did not affect their overall conscious reliance on cross-linguistic similarity. This may help explain why the awareness training did not have an impact on lexical acquisition.

Discussion

Research shows that cognates are better known than noncognates and false cognates are less known than cognates and noncognates (e.g., Otwinowska & Szewczyk, 2019; Puimège & Peters, 2019; Silva & Otwinowska, 2019). Here, we investigated whether this cognate advantage and false cognate disadvantage may stem from contextual word learning (CWL). Over three days, we presented participants with 45 short stories with keywords (cognates, false cognates, and noncognates) embedded in contexts with different degrees of informativeness and occurring 2 to 10 times across the stories. Participants were divided into Experimental and Control groups, and the Experimental group had their awareness of cross-linguistic similarity raised. We measured word learning using a pretest-posttest design while controlling for guessing.

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

Differences in the CWL of noncognates, cognates, and false cognates

Regarding RQ1, the results revealed robust CWL of all three word types. Cognates were learned better than noncognates, and false cognates were learned worse. We discuss these findings below.

Cognates vs. noncognates

Cognates were learned more effectively than noncognates even under strict control for cognate guessing: We considered correct only correct translations with participant-reported full confidence. Such cognate advantage corroborates intentional-learning research using isolated words or sentences in controlled settings (e.g., De Groot & Keijzer, 2000; Elias & Degani, 2021), but contrasts with classroom research (Otwinowska et al., 2020; Rogers et al., 2015). Our findings support Otwinowska et al.'s (2020) hypothesis that the cognate advantage over noncognates emerges during contextual word learning. Our results also support CWL research using audiovisual input (Peters, 2019; Peters & Webb, 2018; Fievez et al., 2020) and the listening and reading of academic texts (Vidal, 2003, 2011). There is thus converging evidence, part of a growing body of research, that the cognate advantage over noncognates may originate from CWL. Importantly, as shown in our alternative analyses in Appendix 13, the higher learning of cognates relative to noncognates was driven by accuracy in translation, not by pretest-to-posttest increase in confidence in translation. Our within-subject design ensured that between-participant variation in self-confidence would not affect results.

False cognates vs. noncognates

Our results showed weaker learning of false cognates than noncognates. This seems expected, as previous findings indicate that false cognates are translated less accurately and are less known than noncognates (Janke & Kolokonte, 2015; Otwinowska & Szewczyk, 2019). Still, these results do not corroborate the results of intentional-learning studies. Marecka et al. (2021) found a learning advantage of false cognates over noncognates when L1-Polish participants learned Polish-like nonwords, similarly to Hirosh and Degani (2021), who investigated the learning of German words via English. On the other hand, Elias and Degani (2021) found no difference between Arabic false cognates and noncognates learned by L1-Hebrew speakers. Therefore, it seems that in intentional learning, false cognates may be acquired better than noncognates if the language of instruction (or participants' L1) and L2 are formally similar. Attention to form may enhance false-cognate learning, in agreement with the Parasitic Model (Ecke, 2015), whereby new L2 forms attach to pre-existing L1 forms. Yet, the classroom study by Otwinowska et al. (2020) found no difference in learning between noncognates and false cognates, even though participants' L1-Polish and L2-English shared the same script. This may be because Otwinowska et al.'s (2020) learning tasks aimed at practicing the meaning of the words, possibly diverting attention from form. Also, classroom exercises necessitated much mental effort, which increased processing and learning of all words. In CWL this should not be the case, because keyword processing would likely be lower.

The current study supports this argument: It appears that in CWL through reading, false cognates are learned worse than noncognates. A similar result for CWL was obtained by Vidal (2011), who reported less learning of deceptively transparent words (false cognates being one of three subtypes) than noncognates. Still, our study seems to

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

be the only CWL research on reading to directly compare the learning of noncognates and false cognates, as opposed to a broader category containing other deceptively transparent words (Vidal, 2011). Possibly, when reading, learners initially link the word form to the wrong meaning, hampering learning. The context where the false cognate appears would need to be highly informative for semantic interference to be overcome and successful semantic inferencing to occur, and for the word to be remembered, as discussed below.

Contextual support and word learning

To answer RQ2, we operationalized contextual support as 1) the number of keyword repetitions and as 2) context informativeness, with more informative contexts providing more clues to the meaning of the word. We found that a higher number of repetitions increased the learning of cognates, noncognates, and false cognates similarly. This is in line with previous research showing that more encounters with novel words in context yield more lexical learning (e.g. Chen & Truscott, 2010; Uchihara et al., 2019; Vidal, 2003, 2011). However, in agreement with Vidal (2011), we expected that cognates would need fewer repetitions than noncognates and false cognates to be learned, but our results did not support this.

It is unclear why this may be the case. It may be that our strict correction for guessing underestimates cognate learning. That is, cognates may have been learned with fewer repetitions, but the correction for guessing masked this effect by making few encounters insufficient to raise participants' confidence to full (7/7 on a Likert scale). This is even more likely considering that the test measurements also included false cognates, which may have undermined participants' confidence in their cognate translations. Further research should verify this.

Also, we found that contextual informativeness predicted learning of all word types, but the effect was small for cognates and noncognates. This corroborates the results of most previous research (e.g., Chen et al., 2017; Elgort et al., 2015; Mulder et al., 2019; Webb, 2008). Regarding cognates, it is possible that repetition alone promotes their learning, irrespective of contextual informativeness. Readers may not need contextual clues to infer the meaning of cognates from context, as the meaning is the same in readers' L1. Upon reading cognates, the learner confirms that the cognate meaning fits the context, and subsequent occurrences of the word serve to confirm its cognateness. As for noncognates, our contexts were overall quite informative. The supportive information comes from the whole text, as the keywords re-occurred across the stories, and from the specific sentences in which the keywords occurred. By comparison, previous L2 studies utilized sentences, not full texts, which often contrasted highly informative to highly uninformative contexts (e.g., Chen et al., 2017; Elgort et al., 2015; Mulder et al., 2019; Webb, 2008; Yi et al., 2022). Such extreme comparisons, beyond the variation typical for natural texts, may be the reason why most of these studies revealed clearer effects for contextual informativeness than what we found.

The effect of context informativeness was significantly more pronounced for false cognates than for the other word types, to the point that false cognates were as probable to be learned as cognates and noncognates in highly informative contexts. It appears that the semantic clues provided by highly informative contexts helped override L1 semantic interference, improving the learning of false cognates. Possibly, although merely speculatively, the conflict between the L1 meaning associated with the form and the one suggested by the context increased the attention directed to false cognates and

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

the urge to infer their meaning. This increased false cognate salience in the input, together with participants' motivation to engage in meaning inferencing—made possible by highly informative contexts—enhanced false-cognate processing and thus learning. Future studies measuring attention focus (e.g., eye tracking; Yi et al., 2022) could verify this.

Cognate awareness-raising and word learning

Answering RQ3, which investigated the impact of cognate awareness-raising on learning, the results showed that the Experimental and Control groups acquired cognates, noncognates, and false cognates to a similar extent, contrary to our expectations. This supports the findings from Otwinowska et al. (2020) that raising awareness of cross-linguistic similarity may not enhance L2 cognate learning. However, the awareness-raising made participants slightly better able, when required, to notice cognates in a novel text. This may indicate participants' increased awareness at the level of noticing, which, however, did not directly translate into learning. Given that both groups reported a similar increase in the use of the Transfer strategy, it is not the awareness raising that affected the intentional reliance on cross-linguistic similarity. One possibility is that the high proportion (thus salience) of cognates and false cognates in the reading increased learners' ability to identify cross-linguistic similarities. Therefore, the Experimental group's small advantage in noticing cognates may not reflect cross-linguistic awareness, but rather a practice effect in identifying cognates in text. Possibly, longer awareness raising might be more successful, but our results suggest that the effect would not be large (see Otwinowska et al., 2020; White & Horst, 2012). Thus, in CWL, participants' approach to cross-linguistic similarity is not easily amenable to explicit awareness raising, so its practical value might be limited.

Limitations

Our study is not without limitations. In the debriefing interviews, about half the participants claimed to have discovered the purpose of the experiment and looked up on average two keywords outside the experimental session. These data points were removed, but knowledge of the experiment's aim may have affected learning rates. This occurred even though we followed best-practice guidelines for this type of research: i.e., we explored a meaning-focused task, did not draw attention to the keywords, and we hid the true purpose of the experiment and the existence ostest. Also, the awareness training was not responsible for this, since a similar number of participants in the Control and in the Experimental groups claimed to have guessed the purpose of the experiment. Therefore, it appears that our post-experiment interviews underscore a possible, albeit often unnoticed, issue with some CWL/incidental-learning research: Participants suspect the purpose of the experiment despite researchers' best efforts to conceal it. Therefore, in future studies, researchers are advised to measure participants' unwarranted exposure to the keywords (see De Vos et al., 2018, for a similar suggestion). Finally, the use of only one test measurement (meaning recall) may also limit the generalizability of our findings, as measurements tapping into different aspects of a word's knowledge may find different results (see Malone, 2018; Webb, 2007, for interesting discussions).

Conclusions

One important finding of this study is that in contextual word learning through reading, under highly controlled conditions, cognates are learned better and false

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

cognates are learned worse than noncognates. This is true despite our strict control for guessing (we accepted as correct only correct translations with full confidence), and irrespective of whether participants undergo cross-linguistic awareness training. This result replicates Otwinowska et al. (2020), who did not find any links between awareness raising and cognate learning in the context of classroom instruction.

Another major conclusion of this study is that false cognates benefited from contextual informativeness significantly more than cognates and noncognates. This may be because readers needed to use the contextual clues of highly informative contexts to overcome the L1 semantic interference of false cognates. It is also possible that this L1-L2 meaning-form conflict made false cognates more salient in the input. This higher salience increased attempts at semantic inferencing, enhancing the effect of context informativeness.

Also, although we expected that cognate learning would occur after fewer repetitions than noncognates and false cognates, we found that the number of keyword occurrences in the written input affected the learning of all three word types to a similar extent. Importantly, however, our alternative models (see Appendix 13) show that our strict correction for guessing did not affect our main conclusions about learning.

Competing interests. The authors declare none.

Data availability statement. All data and materials can be found in Appendices 1–14, openly available on the Open Science Framework at https://osf.io/vfzwb/overview. Also, all raw data and scripts reproducing the analyses can be found online at: https://osf.io/vfzwb/overview.

Acknowledgments. Agnieszka Otwinowska and Jakub Szewczyk contributed equally to this work. The research was supported by grant 2016/21/B/HS6/01129 from the National Science Centre. Poland awarded to Agnieszka Otwinowska-Kasztelanic.

Funding source. This research is part of the project No. 2022/47/P/HS6/02294 within the POLONEZ BIS programme co-funded by the National Science Centre and the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 945339.

References

Amazon Mechanical Turk. (n.d.). Retrieved www.mturk.com

Bordag, D., Gor, K., & Opitz, A. (2022). Refining key concepts of the Ontogenesis Model of the L2 lexical representation. *Bilingualism: Language and Cognition*, *25*(2), 236–241. https://doi.org/10.1017/S1366728921000894

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911. https://doi.org/10.3758/s13428-013-0403-5

Brzeziński, J., Gaul, M., Hornowska, E., Jaworowska, A., Machowski, A., & Zakrzewska, M. (2004). *WAIS-R (PL)—Skala Inteligencji Wechslera dla Dorosłych—Wersja Zrewidowana*. Pracownia Testów Psychologicznych.

Cambridge English Proficiency Test. (n.d.). Retrieved 15 November 2023, from https://www.cambridgeenglish.org/test-your-english/general-english/

Chen, C., & Truscott, J. (2010). The Effects of Repetition and L1 Lexicalization on Incidental Vocabulary Acquisition. *Applied Linguistics*, *31*(5), 693–713. https://doi.org/10.1093/applin/amq031

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

Chen, O., Woolcott, G., & Sweller, J. (2017). *Using cognitive load theory to structure computer-based learning including MOOCs*. Blackwell Science.

Cobb, T. (2026). *Vocabprofile*. Lextutor. http://lextutor.ca

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge University Press.

De Groot, A. (2011). *Language and cognition in bilinguals and multilinguals: An introduction*. Psychology Press.

De Groot, A., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, *50*(1), 1–56. https://doi.org/10.1111/0023-8333.00110

De Vos, J. F., Schriefers, H., & Lemhöfer, K. (2019). Noticing vocabulary holes aids incidental second language word learning: An experimental study. *Bilingualism: Language and Cognition*, *22*(3), 500–515. https://doi.org/10.1017/S1366728918000019

Ecke, P. (2015). Parasitic vocabulary acquisition, cross-linguistic influence, and lexical retrieval in multilinguals. *Bilingualism: Language and Cognition*, *18*(02), 145–162. https://doi.org/10.1017/S1366728913000722

Ecke, P., & Hall, C. J. (2022). Bilingual aspects of the ontogenesis model: Parasitic connections at all levels of representation? *Bilingualism: Language and Cognition*, *25*(2), 208–209. https://doi.org/10.1017/S1366728921000675

Elgort, I., Candry, S., Boutorwick, T. J., Eyckmans, J., & Brysbaert, M. (2018). Contextual word learning with form-focused and meaning-focused elaboration. *Applied Linguistics*, *39*(5), Article 5. https://doi.org/10.1093/applin/amw029

Elgort, I., Perfetti, C. A., Rickles, B., & Stafura, J. Z. (2015). Contextual learning of L2 word meanings: Second language proficiency modulates behavioural and event-related brain potential (ERP) indicators of learning. *Language, Cognition and Neuroscience*, *30*(5), 506–528. https://doi.org/10.1080/23273798.2014.942673

Elias, M., & Degani, T. (2022). Cross-language interactions during novel word learning: The contribution of form similarity and participant characteristics. *Bilingualism: Language and Cognition*, *25*(4), 548–565. https://doi.org/10.1017/S1366728921000857

Fievez, I., Perez, M. M., Cornillie, F., & Desmet, P. (2020). Vocabulary Learning Through Viewing Captioned or Subtitled Videos and the Role of Learner- and Word-Related Factors. *CALICO Journal*, *37*(3), 233–253.

Godfroid, A., Ahn, J., Choi, I., Ballard, L., Cui, Y., Johnston, S., Lee, S., Sarkar, A., & Yoon, H.-J. (2017). Incidental vocabulary learning in a natural reading context: An eye-tracking study. *Bilingualism: Language and Cognition*, *21*(3), 563–584.

Hirosh, Z., & Degani, T. (2021). Novel word learning among bilinguals can be better through the (dominant) first language than through the second language. *Language Learning*, *71*(4), 1044–1084. https://doi.org/10.1111/lang.12457

Hu, H. M., & Nation, P. (2000). What vocabulary size is needed to read unsimplified texts. *Reading in a Foreign Language*, *8*, 689–696.

Janke, V., & Kolokonte, M. (2015). False cognates: The effect of mismatch in morphological complexity on a backward lexical translation task. *Second Language Research*, *31*(2), 137–156. https://doi.org/10.1177/0267658314545836

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, *42*(3), 627–633.

Laufer, B., & Hulstijn, J. (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

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Malone, J. (2018). Incidental vocabulary learning in SLA: Effects of frequency, aural enhancement, and working memory. *Studies in Second Language Acquisition*, *40*(3), 651–675. https://doi.org/10.1017/S0272263117000341

Marecka, M., Szewczyk, J., Otwinowska, A., Durlik, J., Foryś-Nogala, M., Kutyłowska, K., & Wodniecka, Z. (2021). False friends or real friends? False cognates show advantage in word form learning. *Cognition*, *206*, 104477. https://doi.org/10.1016/j.cognition.2020.104477

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

Mulder, E., Van De Ven, M., Segers, E., & Verhoeven, L. (2019). Context, word, and student predictors in second language vocabulary learning. *Applied Psycholinguistics*, *40*(1), 137–166. https://doi.org/10.1017/S0142716418000504

Nation, I. S., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Heinle, Cengage Learning.

Otwinowska, A. (2015). *Cognate vocabulary in language acquisition and use: Attitudes, awareness, activation*. Multilingual Matters.

Otwinowska, A., Foryś-Nogala, M., Kobosko, W., & Szewczyk, J. (2020). Learning orthographic cognates and non-cognates in the classroom: Does awareness of cross-linguistic similarity matter? *Language Learning*, *70*(3), 685–731. https://doi.org/10.1111/lang.12390

Otwinowska, A., & Szewczyk, J. M. (2019). The more similar the better? Factors in learning cognates, false cognates, and non-cognate words. *International Journal of Bilingual Education and Bilingualism*, *22*(8), 974–991.

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly*, *53*(4), 1008–1032. https://doi.org/10.1002/tesq.531

Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, *40*(3), Article 3. https://doi.org/10.1017/S0272263117000407

Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*(4), 878–912. https://doi.org/10.1111/lang.12079

Puimège, E., & Peters, E. (2019). Learners' English Vocabulary Knowledge Prior to Formal Instruction: The Role of Learner-Related and Word-Related Variables. *Language Learning*, *69*(4), 943–977. https://doi.org/10.1111/lang.12364

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, *1*(8), 9.

Rogers, J., Webb, S., & Nakata, T. (2015). Do the cognacy characteristics of loanwords make them more easily learned than noncognates? *Language Teaching Research*, *19*(1), 9–27. https://doi.org/10.1177/1362168814541752

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Silva, B., & Otwinowska, A. (2019). VST as a reliable academic placement tool despite cognate inflation effects. *English for Specific Purposes*, *54*, 35–49. https://doi.org/10.1016/j.esp.2018.12.001

Szewczyk, J. M., & Federmeier, K. D. (2022). Context-based facilitation of semantic access follows both logarithmic and linear functions of stimulus probability. *Journal of Memory and Language*, *123*, 104311. https://doi.org/10.1016/j.jml.2021.104311

Taylor, W. L. (1957). 'Cloze' readability scores as indices of individual differences in comprehension and aptitude. *Journal of Applied Psychology*, *41*, 19–26. https://doi.org/10.1037/h0040591

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

Vidal, K. (2003). Academic listening: A source of vocabulary acquisition? *Applied Linguistics*, *24*(1), 56–89. https://doi.org/10.1093/applin/24.1.56

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition: Incidental vocabulary acquisition. *Language Learning*, *61*(1), 219–258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46–65. https://doi.org/10.1093/applin/aml048

Webb, S. (2008). The effects of context on incidental vocabulary learning. *Reading in a Foreign Language*, *20*(2), 232–245.

White, J. L., & Horst, M. (2012). Cognate awareness-raising in late childhood: Teachable and useful. *Language Awareness*, *21*(1–2), 181–196. https://doi.org/10.1080/09658416.2011.639885

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., … Rush,

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press

A. (2020). Transformers: State-of-the-art natural language processing. In Q. Liu & D. Schlangen (Eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations* (pp. 38–45). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.emnlp-demos.6

Yi, W., Lu, S., & Dekeyser, R. (2022). Orthographic, semantic, and contextual influences on initial processing and learning of novel words during reading: Evidence from eye movements. *Chinese Journal of Applied Linguistics*, *45*(2), 194–219. https://doi.org/10.1515/CJAL-2022-0203

Cite this article: Silva, B., Foryś-Nogala, M., Kutyłowska, K., Szewczyk, J., & Otwinowska, A. (2026). Learning cognates, false cognates, and noncognates contextually through reading: Do contextual support and cross-linguistic awareness matter? *Studies in Second Language Acquisition,* 48: 808–832. https://doi.org/10.1017/S0272263126101831

https://doi.org/10.1017/S0272263126101831 Published online by Cambridge University Press