




cambridge.org/bil

Liat Nativ¹, Yuval Nov², Noam Ordan³, Shuly Wintner¹ and Anat Prior 

Research Article

Cite this article: Nativ, L., Nov, Y., Ordan, N., Wintner, S., & Prior, A. (2024). Do more proficient writers use fewer cognates in L2? A computational approach. *Bilingualism: Language and Cognition*, **27**, 84–94. <https://doi.org/10.1017/S1366728923000482>

Received: 2 June 2022
Revised: 19 April 2023
Accepted: 5 June 2023
First published online: 5 October 2023

Keywords:
proficiency; cognates; written production; computational; proficiency

Corresponding Author:
Anat Prior
Department of Learning Disabilities and Edmond J. Safra Brain Research Center for Learning Disabilities
Faculty of Education, University of Haifa, Haifa, Israel
E-mail: aprior@edu.haifa.ac.il

¹Department of Computer Science, University of Haifa, Haifa, Israel; ²School of Public Health, University of Haifa, Haifa, Israel; ³The Israeli Association of Human Language Technologies, Israel and ⁴Department of Learning Disabilities and Edmond J. Safra Brain Research Center for Learning Disabilities, Faculty of Education, University of Haifa, Haifa, Israel

Abstract

Bilinguals often show evidence of cross language influences, such as facilitation in processing cognates. Here we use computational methods for analyzing spontaneous English texts written by hundreds of speakers of different L1s, at different levels of English proficiency, to investigate writers' preference for using cognates over alternative word choices. We focus on English, since a majority of its lexicon is either of Romance or Germanic origin, allowing an investigation of the preference of speakers of Germanic and Romance L1s towards cognates between their L1 and English. Results show that L2 writers tend to prefer English cognates, and that this tendency is weaker as English proficiency level increases, suggesting diminishing effects of CLI. However, a comparison of the L2 writers with native English writers shows general overuse of cognates only for the Germanic, but not the Romance, L1 speakers, most likely due to the register of argumentative writing.

Introduction



The two languages of bilinguals, who are a majority in the world today (Grosjean & Li, 2013), are not independent of each other (Kroll et al., 2014; Prior, 2014). Specifically, there are influences from speakers' first language (L1) on their second language (L2) (and also vice versa, e.g., Degani et al., 2011), a phenomenon which is termed TRANSFER OR CROSS LANGUAGE INFLUENCE (CLI) (Jarvis & Pavlenko, 2008; Odlin, 1989; van Hell & Tanner, 2012). CLI is evident in various language domains, including phonology, morphology, lexicon and grammar, and is one of the reasons for differences between L1 and L2 speakers of the same language. In fact, these differences are so prominent that even highly advanced L2 speakers can be accurately distinguished from L1 speakers (Bergsma et al., 2012; Goldin et al., 2018; Rabinovich et al., 2016; Tomokiyo & Jones, 2001). In the current study of CLI, we focus on L2 speakers' preference for L2 words that have a cognate in the speakers' L1. Our hypothesis is that CLI, as reflected by lexical choice, is correlated with the speaker's L2 proficiency: more proficient speakers will show lower levels of CLI (Degani et al., 2022). We employ corpus-based computational methods to investigate this hypothesis.

This work focuses on cognates, as a much studied test case for CLI. Cognates are words in different languages that have similar forms and similar meanings, either due to a common ancestor in some protolanguage or via borrowing (for example, *sofa* in English and in Spanish). Due to this cross language similarity, when bilinguals need to process or retrieve a word from the lexicon, cognates are more easily accessed because they are activated in both language systems (e.g., Degani et al., 2018; Dijkstra & Van Heuven, 1998, 2002). Such activation, which is non-selective for language, results in facilitation effects when bilinguals learn or process cognates. Thus, bilinguals are faster to recognize and respond to cognates, relative to non-cognates, when they are presented visually (Dijkstra et al., 2010) or aurally (Woutersen et al., 1995). Cognate facilitation has also been observed using eye tracking: during text reading, bilinguals read cognates faster than non-cognates (Cop et al., 2017; Libben & Titone, 2009). Along similar lines, bilinguals are faster and less error prone when producing cognates, when they are reading or translating words out loud (de Groot, 1992; Schwartz et al., 2007; van Hell & de Groot, 1998), naming pictures (e.g., Costa et al., 2000; Hoshino & Kroll, 2008), or typing words (Muylle et al., 2022).

Here we investigate a somewhat different facet of cognate use in bilinguals: Namely, do bilinguals show preference for using cognates when there is an option to do so? In a series of studies, Prior and colleagues bring evidence to this effect. In an offline single word translation task, Prior et al. (2007) tested moderately proficient bilingual speakers of English and Spanish, and reported that for words that had two or more plausible translations, bilinguals showed a strong preference towards producing a cognate translation, if one existed. In a second study, this preference was also evident in a

© The Author(s), 2023. 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.



  This article has earned badges for transparent research practices: Open Data and Open Materials. For details see the Data Availability Statement.

timed word translation task (Prior et al., 2013). When translating translation-ambiguous words (namely, words with more than a single translation, Schwieter & Prior, 2019), bilinguals once again were more likely to produce the cognate translation if one existed, and were also faster and more accurate when doing so. Finally, this tendency of bilinguals to prefer cognate translations was also evident in professional translators working on full-length texts (Prior et al., 2011). Thus, when bilinguals have an option in word selection, because two words in the target language are synonyms (or close synonyms) of each other, and when one of these words is a cognate with the speakers' L1, bilinguals have a preference for producing the cognate.

Corpus-based computational work has also identified cognate word choice as an important phenomenon shaping the language of L2 speakers. Rabinovich et al. (2018) introduced the L2-Reddit corpus: "a large corpus of highly-advanced, fluent, diverse, nonnative English, with sentence-level annotations of the native language of each author". The resulting dataset included texts written in English by authors with 31 different L1s. Next, they used WordNet and Etymological WordNet (de Melo, 2014) to automatically construct a focus list of synonym sets, such that each set included at least two synonyms with at least two different etymologies. This focus set was further revised to eliminate cultural bias. Based solely on the frequencies of the words in the focus set, and using a hierarchical clustering algorithm, Rabinovich et al. (2018) were able to reconstruct the phylogenetic language tree of the Indo-European language family. In other words, they demonstrated that speakers of different L1s tend to prefer cognates when they write in English, to the extent that it is possible to identify their L1 based mainly on the frequencies of these English words in their texts.

Thus, across both behavioral and computational approaches, there is convincing evidence that when using an L2, bilinguals demonstrate the effects of CLI in their preference for using words that are cognates with their L1. In the current study, we investigate a possible link between L2 proficiency and such cognate preferences. Specifically, we ask whether this preference for cognates is weaker in more proficient L2 speakers. Several previous studies support such a link between proficiency and cognate facilitation effects. For example, Kroll et al. (2002) examined two groups of less and more proficient bilinguals of English (L1) and French (L2) in word naming and word translation tasks. The cognate effect was significant for both groups, but was consistently more prominent for the less proficient bilinguals (see also Poarch & van Hell, 2012). Along similar lines, Rosselli et al. (2014) asked balanced and unbalanced Spanish-English bilinguals to name pictures, denoting cognates and non-cognates, in Spanish and English. The balanced bilinguals demonstrated cognate effects of similar magnitude in the two languages, but the unbalanced bilinguals showed a larger cognate effect when naming in the non-dominant language than in the dominant language (for reviews, see van Hell et al., 2019; van Hell & Tanner, 2012). However, such a link between proficiency and cognate facilitation is not always evident. For example, in a study of visual word processing in bilingual speakers of Arabic and Hebrew, which do not share a script, Degani et al. (2018) again demonstrated robust cognate effects, but these were not modulated by L2 proficiency (see also Prior et al., 2017).

Importantly, these and most other related studies were conducted in a laboratory environment, based on a limited number of participants, native languages, and target words. In contrast, here we use computational analysis to conduct a corpus-based study including thousands of argumentative essays, authored by hundreds of learners with four different L1s, and hundreds of target words. Sampling a more diverse population, who are

responding to a free production task, might be more sensitive for identifying the possible modulating impact of proficiency on cognate effects than lab-based comprehension/judgments tasks.

We hypothesize that the preference for cognates in lexical selection, which is clearly evident in the works reviewed here and in many others, is correlated with L2 proficiency, i.e., that it weakens as the L2 proficiency level increases, becoming more similar to native preference. Ideally, we would put this hypothesis to test by using an extensive corpus, representing multiple levels of speakers from a wide range of native languages, much like the L2-Reddit corpus mentioned above (Rabinovich et al., 2018). Such corpora, however, include very little metadata, and in particular are not tagged for user proficiency. Using common measures, both lexical (Kyle & Crossley, 2015) and syntactic (Lu & Ai, 2015), to assess L2 proficiency on such enormous and noisy data sets is complex, expensive, and also sensitive to spurious factors such as prompt (task) and L1, which might pose a real difficulty in this regard (Lu & Ai, 2015; Weiss, 2017). Therefore, we decided to examine the hypothesis using a "cleaner" corpus, which is tagged both for user's L1 and for their level of L2 proficiency – namely, the TOEFL corpus, a dataset of essays written in English by non-native English speakers wishing to enroll in English-speaking universities (Blanchard et al., 2013; Malmasi et al., 2017). Of relevance, we defined proficiency here rather technically, as scores in the TOEFL language aptitude test, without directly addressing the important question regarding the correspondence between standard aptitude tests and "real" language proficiency (e.g., Wisniewski, 2018).

We test our hypothesis using English texts authored by non-native speakers of English whose L1 is either of Germanic or of Romance origin. This task lends itself particularly well to English as the target language, since despite its Germanic origins, English vocabulary is heavily influenced by Romance languages, mainly French. Historical linguists estimate that about 11,000 words (mainly French and Latin) entered the English lexicon during the Middle English period (Culpeper & Clapham, 1996). The result is that nowadays there are numerous word pairs with different etymology that have approximately the same meaning, such as *speed/velocity*, *start/commence*, where the former is Germanic and the latter is Romance. The Germanic words are often associated with a lower register while their Romance counterparts are considered of a higher register (Franceschi, 2019). Further, native English speaking children learn Germanic words at an earlier age than their Latin-based counterparts (Hernandez et al., 2021). In addition, we also documented the lexical choices of native English speakers to these same word pairs, as evident in a corpus of essays written by native English speakers – namely, LOCNESS (Granger, 1998) – and in frequency distributions in a large sample of English (COW; Schäfer, 2015).

We apply computational methods on the TOEFL texts to quantify the tendency of the L2 writers to use words which have a common origin with their L1. Thus, we investigate how lexical choice, as reflected by the use of cognates, is correlated with L2 proficiency on a large scale, targeting hundreds of cognates that occur in argumentative essays written by thousands of learners with several different L1s. We further compare the tendencies of L2 writers to the tendencies of English L1 writers, to test our hypothesis that with increasing proficiency the patterns of L2 writers will grow more similar to those of L1 writers.

Method

Dataset

The main corpus used in the current study is the TOEFL Corpus (Blanchard et al., 2013; Malmasi et al., 2017), a dataset of essays

written by nonnative English speakers wishing to enroll in English-speaking universities. The essays were evaluated for proficiency by highly skilled annotators, and each was given a grade in the range {*low*, *medium*, *high*}. The corpus consists of 12,100 essays written by native speakers of 11 native languages: 1330 graded *low*, 6568 graded *medium* and the remaining 4202 graded *high*. Each L1 is represented by 1100 essays across all 3 levels. Metadata fields include the author's L1, their proficiency level, and (the index of) the prompt question for the essay (1–8).

In the current investigation, we selected from this corpus nonnative speakers of English whose L1 is of Romance or Germanic origin. The former is represented by Italian, Spanish, and French, whereas the latter only by German. Table 1 shows dataset information by level and by the language and language family of the writers' L1. Evidently, the dataset is not balanced, in more than one aspect. First, it includes significantly more essays written by authors with a Romance L1 compared with German as an L1. Also, in each L1 separately, as well as in the complete dataset, the number of essays per level is unbalanced: the numbers of low-level essays are always smaller. This effect intensifies in the number of words, because the low level essays are also shorter than the medium and high level essays, leading to much smaller text samples for low proficiency writers. A standard solution to this problem is down-sampling: randomly selecting subsets of each group, each in the size of the smallest group. This would have resulted in extremely small samples in our case, hampering our ability to yield meaningful results. We opted instead to retain the unbalanced corpus.

To compare the non-native essays with similarly authored native texts, we used a comparable native-speaker corpus, LOCNESS, the Louvain Corpus of Native English Essays (Granger, 1998). We used a subset of 412 essays written by

A-level native English speakers from British and American universities. The essays in this corpus are much longer than TOEFL essays. Still, the characteristics of the two datasets are as similar as can be: the writers are of similar age, the genre is argumentative writing, and the setting is a test (see Table 2).

Although the LOCNESS dataset enables a relatively fair comparison with TOEFL, it does not necessarily reflect the common frequency distribution of the target words in the English language in general. The corpus consists of a collection of argumentative essays, written as part of a university exam, by a relatively small group of A-level students. These settings dictate a certain writing style and choice of words. To estimate a more ecologically valid frequency distribution of target words in English, we also examined their frequencies based on a very large corpus of diverse English – namely, frequency lists based on COW (CORpora from the Web), a huge collection of linguistically processed web corpora (Schäfer, 2015; Schäfer & Bildhauer, 2012). These frequency lists include the word, its part-of-speech tag, and the number of its occurrences in the corpus. While COW corpora are not claimed to be representative of any specific language variety (in fact, they are known to be biased by many factors, specifically the link structure of the Web), their sheer size makes them good candidates for reflecting “standard” language use, to the extent that such a concept exists.

Ideally, we would like to test our hypothesis considering each individual writer separately. However, since writers vary greatly with respect to their style (including essay length, spelling errors, etc.) such an approach runs the risk of reflecting each writer's personal style rather than their proficiency level and use of cognates. To eliminate noise that might result from such confounding variables, and due to the brevity of text per individual writer, we aggregated all same-grade essays with the same L1 language

Table 1: Text statistics by proficiency level and language family.

	<i>Italian</i>	<i>Spanish</i>	<i>French</i>	Romance	Germanic
Low					
essays	164	79	63	306	15
words	35,952	19,132	13,336	68,420	3614
sentences	2761	1229	1165	5155	351
avg sentence length	13.02	15.57	11.45	13.27	10.3
avg essay length	219.22	242.18	211.68	223.59	240.93
Medium					
essays	623	563	577	1763	412
words	193,497	184,581	186,974	565,052	138,464
sentences	13,627	12,509	15,512	41,648	12,346
avg sentence length	14.2	14.76	12.05	13.57	11.22
avg essay length	310.59	327.85	324.05	320.51	336.08
High					
essays	313	458	460	1231	673
words	112,126	176,408	172,868	461,402	259,397
sentences	7600	12,493	14,138	34,231	21,555
avg sentence length	14.75	14.12	12.23	13.48	12.034
avg essay length	358.23	385.17	375.8	374.82	385.43

Table 2: Text size comparison between TOEFL and LOCNESS.

	TOEFL	LOCNESS
Essays	4,400	412
Words	1,496,349	330,668
Sentences	115,286	18,415
Avg sentence length	12.98	17.96
Avg. essay length	340.08	802.59

family (Romance or Germanic). For example, all *low*-graded essays written by German speaking writers are analyzed as a group named “*Germanic low proficiency*”, while all *medium*-graded essays written by Spanish, Italian or French speaking writers are analyzed as a group named “*Romance medium proficiency*”. Similarly, the LOCNESS essays were also analyzed as a single group named “*native speakers*”.

Target word list

In order to investigate how L2 English writers select specific lexical items during written expression in English, we constructed a list of highly frequent words, synonymous in a manner that captures a common sense of the word. A preliminary list of English synonyms was identified using online resources¹, and also relying on words identified in previous research (Rabinovich et al., 2018). These lists were then manually evaluated by the authors with the goal of identifying synonym sets which included mainly medium to high frequency words, which have a greater probability of being used by L2 writers (Mean word frequency was 57.3 per million, SD=186, based on SUBTLEXus – Brysbaert & New, 2009). We then defined a set of English synonym sets (*synsets*): each synset includes two or more synonyms originating from different language families. In addition, we assigned a part-of-speech (POS) tag to each word, in order to reduce ambiguity. The words in each synset are exclusively of Germanic or Romance origin; we used Wiktionary to determine word etymology. The full list includes 235 synsets, each including at least one word with Germanic origin and at least one word of Romance origin, and 537 words in total. Examples include the nouns {*mistake, error*}, where the former is of Germanic origin and the latter is Romance; or the adjectives {*endless, everlasting, eternal, infinite*}, where the first two are Germanic and the last two are Romance (see Table S1 in the supplementary materials, as well as the online repository, for the complete list).

Note that the target words, identified as described above, do not necessarily have cognates in all the various L1s included in our study, even though etymologically they come from the relevant language group. Many do, of course (e.g., *bloom* has a cognate in German, while *flower* has a cognate in French; similarly, Germanic *full* vs. Romance *complete*), but it is possible that some do not. At any rate, the existence of a few synsets which include words that might not have direct cognates in all four L1s only works against our hypothesis and makes it more difficult to find evidence that supports it, because it would add random noise rather than amplify the signal of “real” cognates that we are after here.

Preprocessing

Ideally, we would like to capture all occurrences of words from the list that retain the sense reflected by their synonym set. For

example, synset 102 includes the nouns {*lift, elevator*}. We would not want to miss the plural form (*lifts, elevators*), but also would not like to include the verb sense of *lift*, since it is not a valid alternative for *elevator*. To address this issue we used *spaCy* (Honnibal et al., 2020) to lemmatize and POS-tag the entire dataset (native and nonnative). Although a consideration of part-of-speech reduces ambiguity, it is far from a perfect solution, because different senses of the words can still exist within the same part of speech (e.g., *lift* as a noun can have the meaning of a ride in addition to that of an elevator).

Procedure

In order to investigate lexical choice in L2 writers of English, we refer to the tendency to use English words with a Germanic origin as *Germanic Tendency* (GT), and to the tendency to choose English words with a Romance origin as *Romance Tendency* (RT).

The Germanic Tendency of authors whose proficiency level is *level* with respect to a synset *s* is defined as the number of occurrences, in all essays of level *level*, of words included in *s* that are of Germanic origin, divided by the total number of occurrences in the same essays of all the words included in *s*. Formally, let *S* be the collection of all synsets considered; for a synset $s \in S$, let *G*(*s*) be the words of Germanic origin included in *s*, and *R*(*s*) be the Romance-originating words in *s*. For a set *W* of words, let #(*W*, *level*) be the total number of occurrences of the words in *W* in essays of proficiency level *level*. We then define:

$$GT(level, s) = \frac{\#(G(s), level)}{\#(G(s), level) + \#(R(s), level)} ; RT(level, s) = 1 - GT(level, s) \tag{1}$$

For example, consider synset 79 – namely, $s_{79} = \{excellent, fantastic, great, wonderful\}$. The Germanic and Romance subsets of s_{79} are $G(s_{79}) = \{great, wonderful\}$ and $R(s_{79}) = \{excellent, fantastic\}$, respectively. Table 3 lists the number of occurrences of the words in s_{79} in the essays of Romance L1 writers, for each proficiency level. Then, for example,

$$\begin{aligned} \#(G(s_{79}), medium) &= 358 + 40 = 398; & \#(R(s_{79}), medium) \\ &= 29 + 35 = 64 \\ GT(medium, s_{79}) &= \frac{398}{398 + 64} \approx 0.861; & RT(medium, s_{79}) \\ &\approx 1 - 0.861 = 0.139. \end{aligned}$$

Table 3: Number of occurrences of the words in synset 79, in essays of Romance L1 writers, from each proficiency level.

Word	Source	Proficiency level		
		Low	Medium	High
excellent	Romance	3	29	13
fantastic	Romance	2	35	12
great	Germanic	25	358	375
wonderful	Germanic	3	40	22

Then, we define the Germanic Tendency of authors whose proficiency level is *level* – namely, $GT(level)$, and similarly the Romance Tendency $RT(level)$, as the (macro) average of $GT(level, s)$ (and, respectively, $RT(level, s)$), across all synsets:

$$\begin{aligned} GT(level) &= \frac{1}{|S|} \sum_{s \in S} GT(level, s); \quad RT(level) \\ &= \frac{1}{|S|} \sum_{s \in S} RT(level, s) \end{aligned} \quad (2)$$

Results

Basic descriptive patterns

We first examined the writers' average tendency to select words from the same language family as their L1, for each critical group (writers whose L1 is Romance, and writers whose L1 is Germanic) separately. That is, we examined the Germanic Tendency of German writers, and the Romance Tendency of Italian, French and Spanish writers. We calculated for each group and for each proficiency level (*low*, *medium*, and *high*) the average of this tendency. To compare the L2 writers' word selection preferences with those of native English speakers (which we assume are not influenced, at the group level, by knowledge of additional languages), we repeated the same calculation based on the essays of LOCNESS. Finally, we used the frequency lists from COW to compute a measure of how words from the target synsets are used in a general purpose, enormous corpus of English, outside the constraints of argumentative essay writing. We expected that higher levels of L2 proficiency in L2 writers would be associated with a weaker Germanic or Romance tendency; additionally, we expected native English speakers to select English words of Germanic origin less often than L2 users with a Germanic L1, and words of Romance origin less often than L2 users with a Romance L1.

Due to the limited size of TOEFL and the lower level of lexical richness among L2 learners in general, some words from the target list appear very few times (or not at all) in the text. This sparsity is most evident in the sample of essays receiving grades of low proficiency². To guarantee robustness in the calculation of Germanic and Romance tendencies, a minimum number of occurrences of the word types in each synset is necessary. We

therefore include here only synsets in which all word types appear at least 3 times in the sample of essays at each level of proficiency³.

German L1

Figure 1 presents the Germanic tendency of low, medium and high proficiency German writers, and the baseline frequency estimates: NATIVE speakers reflected by essays in the LOCNESS dataset, and GENERAL (web-crawled) English, based on word frequencies from COW. The Germanic tendency is computed for 15 synsets whose words occur at least $t=3$ times in essays of each level (a total of 232, 6,190, and 10,864 occurrences for the low, medium and high level essays, respectively). As described above (in equation 2) this was calculated as the macro-average. The error bars show the standard error of the mean; the large variability is probably due to the relatively small number of synsets.

The pattern visible in Figure 1 is consistent with the research hypothesis, as the Germanic tendency decreases for German speaking non-native writers who are more proficient in English. Both LOCNESS native authors and the values from COW reflect lower use of the Germanic alternatives within the target synsets compared with those demonstrated by the L2 writers.

Romance L1s

Figure 2 shows the Romance tendency of Romance L1 speakers, based on 75 synsets whose words occur at least $t=3$ times in essays of each level (a total of 2,345, 20,420 and 17,071 occurrences for the low, medium and high level essays, respectively). As above, the figure presents macro-averages and standard errors of the mean.

The values of the Romance tendencies are overall lower than those of the Germanic tendencies. Values under 0.5 mean that in general, even Romance writers prefer the Germanic alternatives over the Romance ones. Focusing only on the three levels of L2 English speakers, a visual inspection again shows that more proficient writers demonstrate a weaker Romance tendency, which is consistent with the hypothesis.

However, the native speakers in the LOCNESS corpus tend to select Romance alternatives at a higher level than do all of the L2 groups, in contrast to our hypothesis. The Romance tendency value based on COW frequencies is between the low and medium levels of the L2 English speakers⁴. We return to these issues following the statistical analyses.

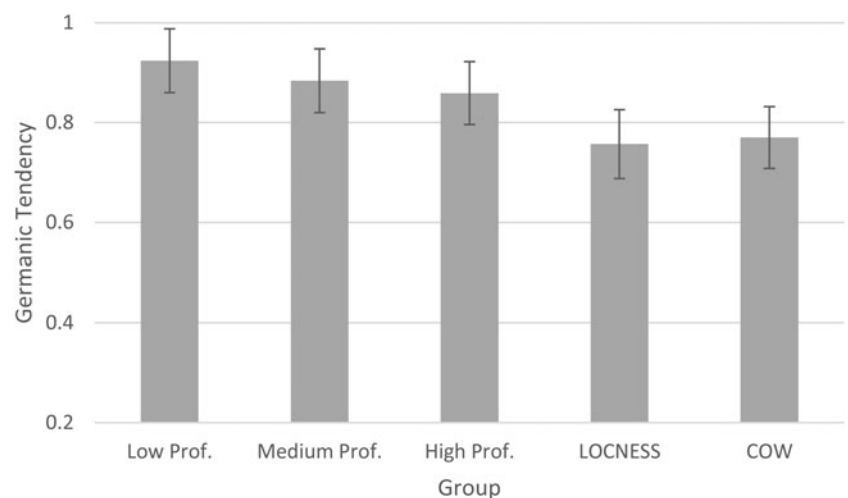


Figure 1: Germanic tendency (GT) of L1 German authors by proficiency and native English authors (Mean, SEM).

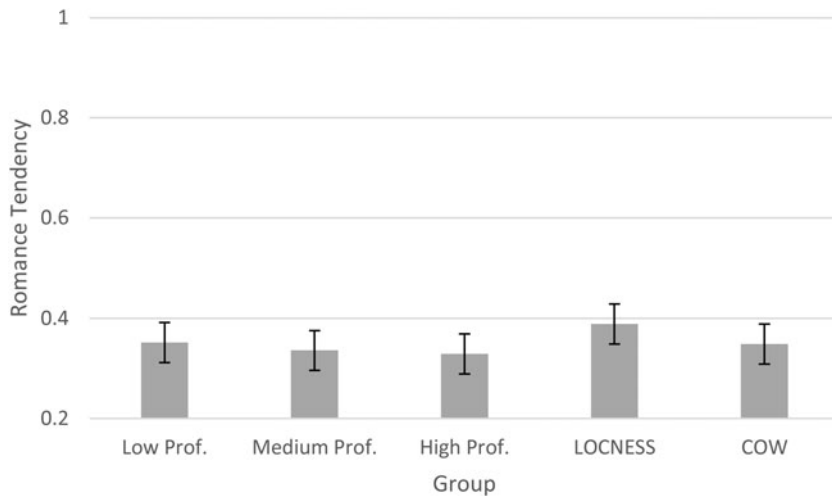


Figure 2: Romance tendency (RT) of Romance authors by proficiency and native English authors (Mean, SEM)

Statistical Analysis: Tendency Permutation

The patterns presented in Figures 1 and 2 support the hypothesis that there is a monotonically decreasing relation between L2 proficiency and the tendency of writers to select L1 cognates. This is reflected in the monotonically decreasing height of the leftmost three columns in both figures. To rule out the possibility that this monotonicity is due to mere chance, we devised and ran a permutation test, tailored to the idiosyncrasies of the data and our hypothesis. The test uses a much larger portion of the data, as it does not require a minimum number of occurrences per synset.

We now describe the test for Romance tendency; the Germanic tendency test is similar, with the obvious changes. The idea is to define a test statistic T , so that for each synset, T is “rewarded” (its value increases) according to the degree at which the writers’ use of words from that synset is consistent with the monotonicity hypothesis. Thus, a large enough value of T supports the hypothesis, and the statistical significance of the test can be derived by comparing the value of T with the distribution of similarly calculated T values, under a suitable random permutation of the data.

Formally, let S_3 be the collection of all synsets whose words appeared in essays written by Romance authors of all three proficiency levels, and let S_2 be defined similarly, for synsets whose words appeared in only two proficiency levels. Synsets whose words appeared in essays of only one of the levels (or not at all) are not included in the analysis.

For a synset $s \in S_3$, let $t(s)$ be the number of inequalities that hold in the hypothesized relation $RT(low, s) > RT(medium, s) > RT(high, s)$, i.e.,

$$t(s) = \begin{cases} 2 & RT(low, s) > RT(medium, s) > RT(high, s) \\ 1 & RT(low, s) > RT(medium, s) \text{ and } RT(medium, s) \leq RT(high, s); \text{ or} \\ & RT(low, s) \leq RT(medium, s) \text{ and } RT(medium, s) > RT(high, s) \\ 0 & \text{otherwise} \end{cases}$$

For a synset $s \in S_2$, define similarly

$$t(s) = \begin{cases} 1 & RT(l_1, s) > RT(l_2, s) \\ 0 & \text{otherwise} \end{cases}$$

where l_1 and l_2 are the two proficiency levels in which the words from s appear, and l_1 is the lower level of the two. We then define the test statistic:

$$T = \sum_{s \in S_2 \cup S_3} t(s)$$

For example, consider again synset 79. The left side of Table 4 lists the Romance tendency corresponding to this synset, of Romance L1 writers from each of the three proficiency levels (computed from the entries of Table 3). In this synset we have $RT(low, s_{79}) > RT(medium, s_{79}) > RT(high, s_{79})$ (because $0.152 > 0.139 > 0.059$), and therefore the contribution of s_{79} to the statistic T is $t(s_{79}) = 2$.

Next, we randomly permuted the Germanic/Romance labels of all words in the dataset. We permute separately the labels in each synset s , in a manner similar to Fisher’s exact test, i.e., while keeping the marginal label counts of s the same as in the original data, both across the proficiency levels (low / medium / high) and across the two etymological sources (Germanic / Romance). See the right side of Table 4 for an example. We generated 10,000 such permutations, and for each permutation i , calculated the corresponding statistic T_i^{perm} .

Under a null hypothesis of no underlying monotonicity in the tendency to choose L1 cognates, the T statistic (based on the original, non-permuted data) is a random observation from the distribution of the $\{T_i^{perm}\}$. However, this null hypothesis is rejected

($p < 0.0001$), as the test statistic was $T = 103$ (based on 177 synsets), higher than all 10,000 T_i^{perm} values. See Figure 2 (right).

We repeated the above procedure to analyze the Germanic tendency, and reached the same result: the test statistic was $T = 77$

Table 4: Number of occurrences and Romance tendency of synset 79, for the original data (L1 Romance authors) and for an example random permutation.

Level	Original Data			Random Permutation		
	Germanic	Romance	Romance Tendency	Germanic	Romance	Romance Tendency
Low	28	5	0.152	31	2	0.0606
Medium	398	64	0.139	407	55	0.119
High	397	25	0.059	385	37	0.0877

(based on 151 synsets), higher than all 10,000 T_i^{perm} values. See the left panel of Figure 2.

We thus robustly established our main hypothesis – namely, that the tendency to use cognates diminishes as L2 proficiency increases. We now set out to evaluate our second question: does L2 writers' tendency to use cognates converge to the levels observed in native speakers? To investigate this, we repeated the described permutation test, this time including the native author group (from the LOCNESS corpus), thus resulting in a maximum of *four* levels of proficiency (low, medium, high, native). We calculated the Germanic and Romance tendencies calculated based on essays written by native writers, and compared the tendency of the high-proficiency non-native writers to that of the native writers. If the tendency of the high proficiency L2 writers was stronger, we added 1 to the value of T_i as described above for the 3 non-native proficiency levels; if not, we added nothing to T_i . Finally, we again created 10,000 random permutations of T_i , including the native writers. Figure 3 shows the histogram of random T_i values for Germanic and Romance tendencies.

In the case of German and native writers, the observed T value (177), based on 183 synsets, is above the maximum value obtained over all 10,000 random permutations (162). This finding supports our hypothesis that the tendency of writers with German L1 to use words of Germanic origin decreases as their proficiency improves and grows more similar to that of native English speakers (Figure 3, left panel).

However, the same pattern is not observed in writers with Romance L1 backgrounds. Here, the observed T value achieved after including the native English writers is 178, based on 196 synsets. Figure 3 (right panel) shows the histogram of the random T values, and the observed T value based on the original texts is in the lower range of the random distribution. This means that when

native English writers are included in the analysis, we no longer have evidence of convergence of non-natives to native writers.

Discussion

We investigated how traces of speakers' L1 can be detected in their L2 lexical selections in production, and in particular through the choice of cognates. Thus, we go beyond demonstrations of cognate facilitation in single word production (e.g., de Groot, 1992; Prior et al., 2007), and comprehension (e.g., Libben & Titone, 2009), and add to the growing literature showing that L2 users have a preference for using cognates when producing written text under natural conditions (Prior et al., 2011; Rabinovich et al., 2018).

Further, we also tested the hypothesis that the tendency to prefer cognates in L2 production weakens as writers' L2 proficiency increases. Some previous research has reported evidence for such an association between cognate facilitation and L2 proficiency (Kroll et al., 2002; Rosselli et al., 2014), but others have not found cognate effects to be modulated by L2 proficiency (Degani et al., 2018; Prior et al., 2017).

Here we offer a different perspective for examining this question, as we extend our investigation beyond the somewhat limited settings of laboratory experiments. Instead of analyzing the responses of a small number of participants requested to complete well controlled tasks, we used computational methods to process spontaneous productions of hundreds of writers. From this large and rich dataset, we calculated writers' tendency to prefer words which share etymology with their L1, and examine whether this tendency is modulated by their L2 proficiency. We also compared the tendencies of L2 writers to select alternatives from the two different etymological sources with those of L1 writers, finding

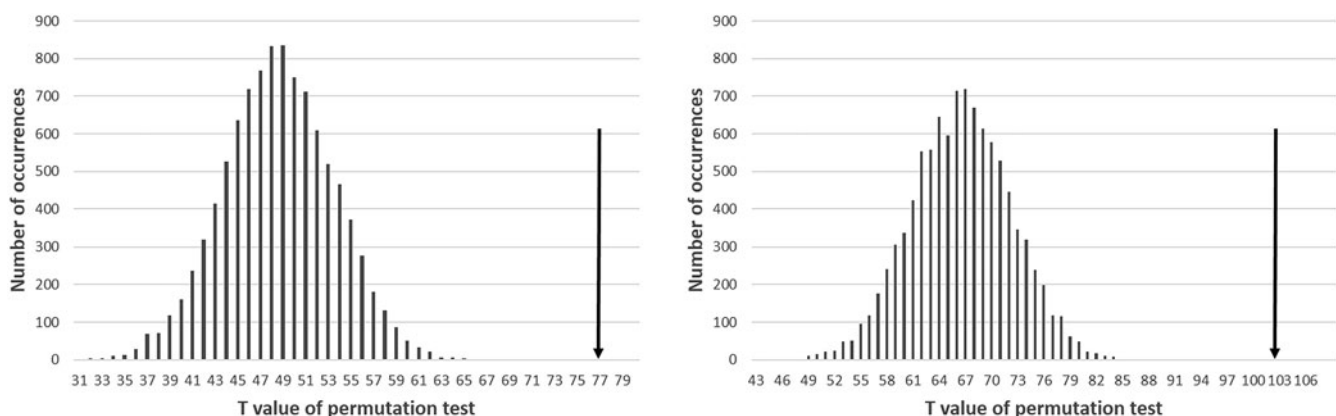


Figure 3: Histograms of the T_i^{perm} values, calculated from random permutation of the data, for Germanic (left) and Romance (right) tendencies. The arrows indicate the T values calculated from the original, non-permuted dataset.

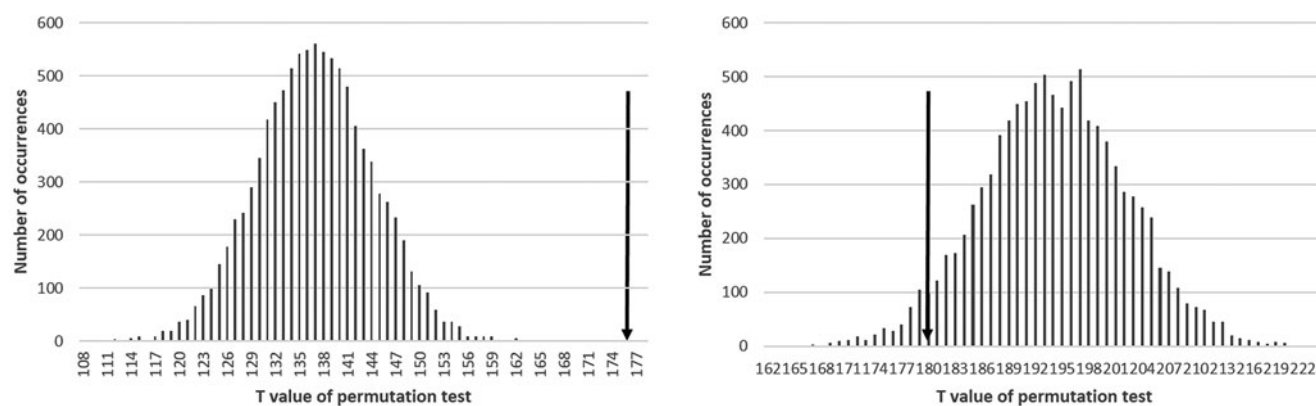


Figure 4: Histogram of random T_i^{perm} values, representing Germanic (Left) and Romance (right) tendencies, when including data based on native author essays in the LOCNESS dataset. The arrows represent T values calculated with the original dataset.

partial support for the hypothesis that with increasing L2 proficiency, L2 writers' lexical selections become more similar to those of L1 authors, though differences were found between the two non-native groups.

For the relatively small Germanic non-native group, consisting exclusively of German L1 writers, the results convincingly supported the hypothesis: the tendency to prefer words of Germanic over Romance source (the GERMANIC TENDENCY) was highest with the least proficient writers, and there was a significant decline in this tendency with increased proficiency. Compared to native English speakers, German L2 writers overuse the Germanic alternatives.

The Romance group included L1 speakers of French, Italian and Spanish. Analysis of the tendency to use Romance words was not fully consistent with the hypothesis. When examining only the L2 speakers, we found that the tendency to use Romance alternatives declined with increased proficiency, significantly more than would be expected by chance. In contrast, the Romance tendency was higher among L1 English speakers than among L2 English speakers at all three proficiency levels, in the genre of argumentative essays examined here. Thus, on the one hand, the comparison among the proficiency levels within the L2 writers supported our hypothesis that lower proficiency writers prefer English words of Romance origin; but the comparison between the L2 and the L1 writers did not support the hypothesis that the L2 writers will overall have a stronger preference for Romance source words than L1 writers.

We ascribe this unexpected finding, that L1 writers used more Romance alternatives than L2 writers whose native language is Romance, to the effects of register within the English language. Thus, the results show that the Germanic alternatives were used more than the Romance alternatives, for all groups included in the study. We suggest that the low usage rates of words with Romance origins can be partially explained by the higher register and lower frequency of English words from Romance origin (Bar-Ilan & Berman, 2007; Franceschi, 2019; Levin & Novak, 1991), and by the fact that they are learned later in life (Hernandez et al., 2021) compared to words of Germanic origin. The native speakers, who most likely have wider English vocabulary knowledge by virtue of greater exposure to the language, can more easily access such less frequent words. Further, recall that the LOCNESS corpus is a collection of essays written during academic exams, a setting in which writers naturally aim at selecting the higher-register, more formal words. However, this explanation

contradicts the stronger Romance tendency of low-proficiency L2 writers, compared with the medium- and even more so with the high-proficiency L2 authors. Going back to the hypothesis, we speculate that the less proficient L2 writers use the higher register Romance alternatives more often, not by virtue of their vocabulary size in English, but rather due to CLI from their L1, and due to their limited exposure to English. The current results do not allow us to directly test this possibility, and we hope that future research may shed more light on this issue.

There are several possible explanations for the finding that higher proficiency in L2 results in reduced CLI – namely, a weaker preference for cognates. First, because they are shared across languages, both L1 and L2 exposure contribute to the lexical frequency of cognates. As frequency distributions stabilize with increased L2 exposure, this impact of the frequency “boost” coming from L1 becomes smaller, because effects of frequency on lexical access are logarithmic in nature. Namely, frequency effects are sizeable at the low end of the distribution, but become smaller at high frequencies (Brybaert et al., 2017; Diependaele et al., 2013; Kuperman & Van Dyke, 2013; Mor & Prior, 2020). Second, in a free writing task, participants have time to consider and monitor their word choice, with less immediate pressure than in many psycholinguistic tasks. Therefore, increased proficiency may lead to a subtler awareness of the appropriateness of the different lexical options in a way that more closely approximates that of L1 speakers. Finally, higher L2 proficiency may indicate enhanced language control, and specifically the ability of bilinguals to manage activation of the non-target language (Abutalebi & Green, 2007; Bonfieni et al., 2019; Costa & Santesteban, 2004; Declerck et al., 2020), i.e., the activation of L1 during L2 use. This again would lead to reduced CLI among more proficient L2 users. These effects are not mutually exclusive, and might be operating in concert to influence the final outcome.

Overall, the current results mostly show reduced CLI in lexical choice, exemplified by a preference for cognates, with increasing L2 proficiency, but previous behavioral studies did not always find L2 proficiency to modulate cognate effects (Degani et al., 2018; Prior et al., 2017). One possible reason for this difference is the current approach, of using large corpora and computational tools, for testing a psycholinguistic hypothesis. By using computational methods, we are able to analyze spontaneous texts written by both L1 and L2 writers, and combine other resources such as frequency lists that are based on very large corpora. Unlike traditional psycholinguistic experiments, we have very limited

information on the individual writers. On the other hand, we are able to dramatically increase the number of participants (writers), and to include more L1s. The broad dataset, along with the more natural manner in which the writers express themselves, compared with a lab experiment, are most likely the cause for the differences found regarding the impact of proficiency on cognate effects.

Limitations and future research

The current study used much more extensive datasets compared to experimental behavioral studies. However, the datasets used here are still considered to be relatively small. The limited size of the dataset forced us, among other things, to go down from the level of the single writer and examine the research hypothesis at a lower resolution, aggregating texts written by different writers graded at the same proficiency level. Testing the hypothesis on larger corpora, where the amount of text contributed by each writer is much larger, will allow analysis of texts written by individuals. If, in addition, corpora will represent multiple levels of speaker proficiency from a wide range of native languages, the insights from this work can be taken further. Specifically, it would be interesting to repeat the experiment with other Germanic languages on a larger scale, and possibly with a wider range of L2 proficiency levels. Unfortunately, such corpora, that are tagged for L1 as well as for L2 proficiency level, especially when it comes to advanced writers rather than learners, are almost non-existent. Building such a dataset would be a great foundation for future research aiming to answer research questions related to the one in this study and others. Alternatively, developing a measure of L2 proficiency that is automatic, reliable, accurate and easy to calculate based on a given text sample, will achieve the same goal, and is likely to be useful for other purposes as well.

Finally, the target word set selected in the current study was based only on etymological information, and might have included English words that do not have direct cognates in some of the L1s. The question of how cognates are identified and defined, and specifically the relative weight of historical linguistic considerations vs. current overlap in form and meaning across languages, is still under debate (e.g., Batsuren et al., 2022). In the current study, the inclusion of English words that do not have a cognate in one or more of the L1s would operate against our ability to find a meaningful signal in the data, and thus does not impede us from reaching meaningful conclusions. However, future research might test different approaches to defining and selecting cognate synonym sets.

Conclusion

We demonstrated robust cognate effects in spontaneous L2 written production, specifically in the lexical choices made by writers. This finding significantly expands our understanding of the dynamics of CLI in a domain of language use that has received limited attention in psycholinguistic research. We further demonstrated that the effects of CLI diminish with increased L2 proficiency, adding important empirical evidence on naturalistic bilingual language use. Specifically, by adopting a computational approach and a large data set, we demonstrated an important finding on the impact of proficiency on CLI, though we cannot at this stage offer a definitive description of the underlying cognitive and linguistic mechanisms, which are ripe for future psycholinguistic investigation. We see the current research, therefore, as

an example of how complementary methodologies from psycholinguistics and natural language processing can lead to fruitful generation and testing of hypotheses, to advance our understanding of CLI in bilingual language processing.

Acknowledgements. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) grant 398186468 and by the Data Science Research Center at the University of Haifa. The authors thank Anke Luedeling, Sarah Schneider, Dominique Bobeck and Chen Gafni for advice and fruitful discussions. The authors also thank Dr. Laura Muscalu and an anonymous reviewer for helpful comments.

Supplementary Material. For supplementary material accompanying this paper, visit <https://doi.org/10.1017/S1366728923000482>

Table S1, full list of target synsets

Competing interests. The authors declare none.

Data availability statement. The data and code that support the findings of this study are openly available in OSF at https://osf.io/3rfx7/?view_only=8c8fb89451a14777ad78f426df8da60b

Notes

- 1 For example, <https://www.englisch-hilfen.de/en/words/synonyms.htm>
- 2 This sparsity made us reluctant to conduct traditional statistical analyses (such as parametric or non-parametric analyses of variance) on these data. Instead, we opted to use a permutation test, utilizing all of the available data, which is presented in the next section.
- 3 Results when using thresholds of 1 and 5 were very similar to those reported here.
- 4 These patterns remained consistent when each of the Romance L1 languages (namely, Spanish, Italian and French) was examined separately, see online repository https://osf.io/3rfx7/?view_only=8c8fb89451a14777ad78f426df8da60b.

References

- Abutalebi, J., & Green, D. (2007). Bilingual language production: The neuro-cognition of language representation and control. *Journal of Neurolinguistics*, 20(3), 242–275. <https://doi.org/10.1016/j.jneuroling.2006.10.003>
- Bar-Ilan, L., & Berman, R. A. (2007). Developing register differentiation: The Latinate-Germanic divide in English. *Linguistics*, 45(1), 1–35. <https://doi.org/10.1515/LING.2007.001>
- Batsuren, K., Bella, G., & Giunchiglia, F. (2022). A large and evolving cognate database. *Language Resources and Evaluation*, 56, 165–189. <https://doi.org/10.1007/s10579-021-09544-6>
- Bergsma, S., Post, M., & Yarowsky, D. (2012). Stylometric analysis of scientific articles. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 327–337. Association for Computational Linguistics.
- Blanchard, D., Tetreault, J., Higgins, D., Cahill, A., & Chodorow, M. (2013). TOEFL11: A corpus of non-native English. *ETS Research Report Series*, (2), i–15.
- Bonfieni, M., Branigan, H. P., Pickering, M. J., & Sorace, A. (2019). Language experience modulates bilingual language control: The effect of proficiency, age of acquisition, and exposure on language switching. *Acta Psychologica*, 193, 160–170. <https://doi.org/10.1016/j.actpsy.2018.11.004>
- 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.
- Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, 20(3), 530–548. <https://doi.org/10.1017/S1366728916000353>

- Cop, U., Dirix, N., van Assche, E., Drieghe, D., & Duyck, W. (2017). Reading a book in one or two languages? An eye movement study of cognate facilitation in L1 and L2 reading. *Bilingualism: Language and Cognition*, 20(4), 747–769. <https://doi.org/10.1017/S1366728916000213>
- Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of Memory and Language*, 50(4), 491–511. <https://doi.org/10.1016/j.jml.2004.02.002>
- Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26(5), 1283–1296. <https://doi.org/10.1037/0278-7393.26.5.1283>
- Culpeper, J., & Clapham, P. (1996). The borrowing of Classical and Romance words into English: A study based on the electronic oxford English dictionary. *International Journal of Corpus Linguistics*, 1(2), 199–218.
- Declerck, M., Kleinman, D., & Gollan, T. H. (2020). Which bilinguals reverse language dominance and why? *Cognition*, 204, 104384. <https://doi.org/10.1016/j.cognition.2020.104384>
- Degani, T., Prior, A., & Tokowicz, N. (2011). Bidirectional transfer: The effect of sharing a translation. *Journal of Cognitive Psychology*, 23, 18–28.
- Degani, T., Prior, A., & Hajajra, W. (2018). Cross-language semantic influences in different script bilinguals. *Bilingualism: Language and Cognition*, 21(4), 782–804. [10.1017/S1366728917000311](https://doi.org/10.1017/S1366728917000311)
- Degani, T., Prior, A., & Wodniecka, Z. (2022). Modulators of cross-language interference in learning and processing. *Frontiers in Psychology*, <https://www.frontiersin.org/articles/10.3389/fpsyg.2022.898793/full>
- de Groot, A. M. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 1001–1018. <https://doi.org/10.1037/0278-7393.18.5.1001>
- de Melo, G. (2014). Etymological Wordnet: Tracing the history of words. In *Proceedings of the 9th Language Resources and Evaluation Conference (LREC 2014)*, Paris, France
- Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: a lexical entrenchment account. *Quarterly journal of experimental psychology*, 66(5), 843–863. <https://doi.org/10.1080/17470218.2012.720994>
- Dijkstra, T., & Van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In Grainger, J., & Jacobs, A.M. (eds.), *Localist Connectionist Approaches to Human Cognition*, pp. 189–225. Mahwah, NJ: Lawrence Erlbaum Associates.
- Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, 5, 175–197.
- Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, 62, 284–301.
- Franceschi, D. (2019). Anglo-Saxon and Latinate Synonyms: The Case of Speed vs. Velocity. *International Journal of English Linguistics*, 9(6), 356, [10.5539/ijel.v9n6p356](https://doi.org/10.5539/ijel.v9n6p356).
- Goldin, G., Rabinovich, E., & Wintner, S. (2018). Native language identification with user generated content. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, 3591–3601. Association for Computational Linguistics, 2018. <http://aclweb.org/anthology/D18-1395>
- Grainger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. *Learner English on Computer*, pages 3–18. Routledge.
- Grosjean, F., & Li, P. (2013). *The Psycholinguistics of Bilingualism*. Wiley-Blackwell.
- Hernandez, A. E., Ronderos, J., Bodet, J. P., Clausenius-Kalman, H., Nguyen, M. V., & Bunta, F. (2021). German in childhood and Latin in adolescence: On the bidialectal nature of lexical access in English. *Humanities and Social Sciences Communications*, 8, 1–12. [10.31219/osf.io/2fjtg](https://doi.org/10.31219/osf.io/2fjtg)
- Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial strength Natural Language Processing in Python, <https://doi.org/10.25205281/zenodo.1212303>
- Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: does cross-language activation survive a change of script? *Cognition*, 106(1), 501–511. <https://doi.org/10.1016/j.cognition.2007.02.001>
- Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic Influence in Language and Cognition*, Routledge.
- Kroll, J. F., Michael, E. B., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, 18, 137–171. <https://doi.org/10.1191/0267658302sr2010a>
- Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate language, cognition, and the brain. *Current Directions in Psychological Science* 23(3), 159–163.
- Kuperman, V., & Van Dyke, J. A. (2013). Reassessing word frequency as a determinant of word recognition for skilled and unskilled readers. *Journal of Experimental Psychology: Human Perception and Performance*, 39(3), 802–823. <https://doi.org/10.1037/a0030859>
- Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49, 757–786. <https://doi.org/10.1002/tesq.194>
- Levin, H., & Novak, M. (1991). Frequencies of Latinate and Germanic words in English as determinants of formality. *Discourse Processes*, 14(3), 389–398. <https://doi.org/10.1080/01638539109544792>
- Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(2), 381–390. <https://doi.org/10.1037/a0014875>
- Lu, X., & Ai, H. (2015). Syntactic complexity in college-level English writing: Differences among writers with diverse L1 backgrounds. *Journal of Second Language Writing*, 29, 16–27.
- Malmasi, S., Evanini, K., Cahill, A., Tetreault, J., Pugh, R., Hamill, C., Napolitano, D., & Qian, Y. (2017). A Report on the 2017 Native Language Identification Shared Task. In *Proceedings of the 12th Workshop on Innovative Use of NLP for Building Educational Applications*, 62–75. <http://aclweb.org/anthology/W17-5007>.
- Mor, B., & Prior, A. (2020). Individual differences in L2 frequency effects in different script bilinguals. *International Journal of Bilingualism*, 24(4), 672–690. <https://doi.org/10.1177/1367006919876356>
- Muylle, M., Van Assche, E., & Hartsuiker, R. (2022). Comparing the cognate effect in spoken and written second language word production. *Bilingualism: Language and Cognition*, 25(1), 93–107. [doi:10.1017/S1366728921000444](https://doi.org/10.1017/S1366728921000444)
- Odlin, T. (1989). *Language Transfer: Cross-linguistic influences in language learning*. Cambridge University Press.
- Poarch, G. J., & van Hell, J. G. (2012). Cross-language activation in children's speech production: Evidence from second language learners, bilinguals, and trilinguals. *Journal of Experimental Child Psychology*, 111, 419–438.
- Prior, A. (2014). Bilingualism: Interactions between languages. In P. Brook & V. Kempe (Eds) *Encyclopedia of Language Development*. Sage Reference. <http://dx.doi.org/10.4135/9781483346441>
- Prior, A., MacWhinney, B., & Kroll, J. F. (2007). Translation norms for English and Spanish: the role of lexical variables, word class, and L2 proficiency in negotiating translation ambiguity. *Behavior Research Methods*, 39(4), 1029–1038. <https://doi.org/10.3758/bf03193001>
- Prior, A., Wintner, S., Macwhinney, B., & Lavie, A. (2011). Translation ambiguity in and out of context. *Applied Psycholinguistics*, 32(1), 93–111. [10.1017/S0142716410000305](https://doi.org/10.1017/S0142716410000305)
- Prior, A., Kroll, J., & Macwhinney, B. (2013). Translation ambiguity but not word class predicts translation performance. *Bilingualism: Language and Cognition*, 16(2), 458–474. [10.1017/S1366728912000272](https://doi.org/10.1017/S1366728912000272)
- Prior, A., Degani, T., Awawdy, S., Yassin, R., & Korem, N. (2017). Is susceptibility to cross-language interference domain specific? *Cognition*, 165, 10–25. <https://doi.org/10.1016/j.cognition.2017.04.006>
- Rabinovich, E., Nisioi, S., Ordan, N., & Wintner, S. (2016). On the similarities between native, non-native and translated Texts. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL*, 1870–1881. <http://aclweb.org/anthology/P16/P16-1176.pdf>
- Rabinovich, E., Tsvetkov, Y., & Wintner, S. (2018). Native language cognate effects on second language lexical choice. *Transactions of the Association for Computational Linguistics*, 6, 329–342. <https://transacl.org/ojs/index.php/tacl/article/view/1403>
- Rosselli, M., Ardila, A., Jurado, M. B., & Salvatierra, J. L. (2014). Cognate facilitation effect in balanced and non-balanced Spanish–English bilinguals using the Boston Naming Test. *International Journal of Bilingualism*, 18(6), 649–662. [10.1177/1367006912466313](https://doi.org/10.1177/1367006912466313)

- Schäfer, R. (2015). Processing and querying large web corpora with the COW14 architecture. In *Proceedings of the 3rd Workshop on Challenges in the Management of Large Corpora (CMLC-3), Lancaster, 20 July 2015* (pp. 28-34). Institut für Deutsche Sprache, <http://rolandschaefer.net/?p=749>.
- Schäfer, R., & Bildhauer, F. (2012). Building large corpora from the web using a new efficient tool chain. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)* (pp. 486-493), <http://rolandschaefer.net/?p=70>.
- Schwartz, A. I., Kroll, J. F., & Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes*, 22(1), 106-129. <https://doi.org/10.1080/01690960500463920>
- Schwietzer, J., & Prior, A. (2019). Translation ambiguity. In R. Heredia & A. Cieśllicka (Eds.), *Bilingual Lexical Ambiguity Resolution*. New York, NY: Cambridge University Press, pp. 96-125.
- Tomokiyo, L. M., & Jones, R. (2001). You're not from 'round here, are you? Naive Bayes detection of non-native utterances. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics*.
- van Hell, J. G., & de Groot, A. M. B. (1998). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition*, 1(3), 193-211. <https://doi.org/10.1017/S1366728998000352>
- van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning*, 62, 148-171. <https://doi.org/10.1111/j.1467-9922.2012.00710.x>
- van Hell, J. G., Donnelly Adams, K., & Abdollahi, F. (2019). Individual variation in bilingual lexical processing: The impact of second language proficiency and executive function on cross-language activation. In J. Darquennes, J. Salmons and W. Vandebussche (eds) *Language Contact. An International Handbook*, pp 210-222. De Gruyter Mouton. doi:10.1515/9783110435351-018. URL.
- Weiss, Z. (2017). Using measures of linguistic complexity to assess German L2 proficiency in learner corpora under consideration of task-effects. *Unpublished MA thesis*. www.sfs.uni-tuebingen.de/~zweiss.
- Wisniewski, K. (2018). The empirical validity of the Common European Framework of Reference scales. An exemplary study for the vocabulary and fluency scales in a language testing context. *Applied Linguistics*, 39(6), 933-959. <https://doi.org/10.1093/applin/amw057>.
- Woutersen, M., de Bot, K., & Weltens, B. (1995). The bilingual lexicon: Modality effects in processing. *Journal of Psycholinguistic Research*, 24, 289-298.