

First Language Test Bias? Comparing French-Speaking and Polish-Speaking Participants' Performance on the Peabody Picture Vocabulary Test

Justyna Leśniewska, François Pichette, and Sébastien Béland

Abstract: Cognates are known to facilitate second language acquisition and use, as learners tend to assign to a new L2 word the meaning of a similar L1 word. Consequently, for L2 tests that rely largely on lexical items, performance may prove inflated for examinees whose L1 shares many cognates with the language being tested. This article examines the possibility of L1 bias on the Peabody Picture Vocabulary Test (PPVT), a well-established measure of receptive vocabulary knowledge in English. To investigate if performance on the PPVT is affected by cognates, we tested 293 speakers of French and 150 speakers of Polish, since those two languages differ markedly in the number of cognates they share with English. After demonstrating that both groups yield clearly distinct response patterns, descriptive and multivariate statistics confirmed that cognate items enhance test performance: the items with the highest score difference in favour of a language group overwhelmingly consist of cognates for that group only. Mantel-Haenszel and logistic regression show that items that are cognates for one of the two groups are more likely to show differential item functioning than the average items. The results suggest that scores on L2 vocabulary-based tests could be biased by the presence of cognates with the examinee's first language.

Keywords: cognates, differential item functioning, language test bias, L1 bias, Peabody Picture Vocabulary Test

Résumé : Les congénères interlinguaux sont connues pour faciliter l'acquisition et l'usage d'une langue seconde (L2), les apprenants ayant tendance à attribuer à un mot nouveau dans une langue seconde le sens d'un mot semblable de la langue première (L1). Il s'ensuit que, dans les tests en L2 qui reposent en grande partie sur les éléments lexicaux, la performance des étudiants interrogés dont la L1 présente de nombreuses accointances avec la langue évaluée risque d'être surestimée. Les auteurs examinent la possibilité d'une distorsion liée à la L1 selon l'échelle de vocabulaire en images Peabody (PPVT), une mesure du vocabulaire compris en anglais qui a fait ses preuves. Afin de déterminer si les accointances influent sur la performance évaluée selon la PPVT, les auteurs testent 293 locuteurs du français et 150 locuteurs du

polonais, ces deux langues présentant des différences marquées au chapitre du nombre d'accointances avec l'anglais. Une fois démontré que les deux groupes affichent des profils de résultats clairement distincts, des statistiques descriptives et multivariées confirment que l'existence d'accointances améliore la performance au test : les éléments présentant l'écart le plus important dans les résultats obtenus en faveur d'un groupe linguistique donné consistent en très grande majorité en accointances qui concernent uniquement ce groupe. La méthode de Mantel-Haenszel et la régression logistique révèlent que les éléments qui sont des accointances pour l'un des deux groupes sont davantage susceptibles d'afficher un fonctionnement différencié que les éléments moyens. Les constatations des auteurs semblent indiquer que les résultats des tests en L2 basés sur le vocabulaire pourraient être altérés par la présence d'accointances avec la langue première de l'étudiant évalué.

Mots clés : congénères interlinguaux, distorsion de test linguistique, distorsion liée à la L1, fonctionnement différentiel des items, test de vocabulaire en images Peabody

The impact of the first language (L1) appears to be the most widely documented and acknowledged phenomenon in second language acquisition. Most, if not all, second language (L2) models account for the influence of the L1 on all aspects of L2 acquisition and use (Ortega, 2009; VanPatten & Williams, 2008; Whong, 2006). Although the influence of the L1 on the acquisition and use of the L2 is the most extensively studied type of cross-linguistic influence, it is now a generally recognized fact that all of a multilingual person's languages influence one another in all directions and in many different ways (Bardel, 2015). Therefore, the general concept of cross-linguistic influence is best defined as "the influence of a person's knowledge of one language on that person's knowledge and use of another language" (Jarvis & Pavlenko, 2007, p. 1).

The focus of this study is on one type of cross-linguistic influence, namely the facilitating effect of L1/L2 cognates on performance on measures of L2 vocabulary knowledge. More specifically, this paper investigates the possible occurrence of L1 bias on the Peabody Picture Vocabulary Test (PPVT), a well-known test of vocabulary knowledge in English, in situations where the test is administered to non-native speakers of English. Since the PPVT is a measure of passive knowledge of individual vocabulary items presented without any context, this means that the topic under investigation is the impact of cognates on the receptive processing of individual lexical items in the L2.

Language similarity and test bias

Fairness toward all examinees has been a central preoccupation in language tests for several decades (Cole & Zieky, 2001). Among the types

of bias a language test can show, some items can be easier for speakers of a certain language than for speakers of another. Since languages differ in how many cognates they share with a specific language, this opens the possibility of particular groups of L1 speakers being favoured on a language test, namely those whose L1 shares some similarities with the language being tested. An indication of possible bias can be obtained by checking if a test displays differential item functioning (DIF), that is, when two groups who are assumed to have the same ability show a different probability of giving a correct response to one or more test items.

Chen and Henning (1985) explored DIF for an English test and identified four items, consisting of language cognates, which favoured speakers of Spanish over speakers of Chinese. In a later DIF study with participants of the same language groups, Sasaki (1991) also showed that some items could favour a language group to the detriment of another. In a more recent study of 184 Korean and 146 Japanese ESL students, Stoeckel and Bennett (2013) showed that the test items that consisted of loanwords favoured the speakers of the language that shared those words with English.

Cognates

Cognates are commonly understood to be words in two different languages that share some similarity in form and meaning. The form of some words may be similar in two languages for two main reasons. First, similarity may be due to the fact that the languages are typologically close, in which case the two words evolved from a single word in a past language that gave rise to the two. The other source of similarity is borrowing, which may occur between any pair of languages, regardless of typological distance. Borrowings are not always direct; they may occur through the medium of one or more languages. For example, the word *candy* went from Persian to French before entering the English language.

These two reasons for cross-linguistic lexical similarity are key elements in the various definitions of “cognate.” One approach to defining this term, often adopted in historical linguistics, is to define cognates as related by descent from the same ancestral language. The words’ origin is the key feature, and they do not necessarily have to be orthographically or phonologically similar (Matthews, 2014). As Otwinowska (2015) notes, in the case of some groups of words that share a common Proto-Indo-European origin, such as the English *mill*, German *Mühle*, French *moulin*, and Polish *młyn*, the common origin is hard to detect on the basis of word form. The other approach, more common in psycholinguistics and in second language acquisition

studies, is to define cognates as words that are similar or identical in two languages, regardless of whether they come from the same source, or have been borrowed from some other language (Dijkstra, Miwa, Brummelhuis, Sappelli, & Baayen, 2010; Yudes, Macizo, & Bajo, 2010). Indeed, it makes little sense to consider the genetic status of cognates in studies of second language acquisition, because from a learner-oriented perspective it makes no difference what the source of similarity is. This article adopts the latter approach and defines cognates as words that share phonological and/or orthographic similarity, regardless of etymological status. It should be noted here that most definitions of cognates either explicitly restrict this category to cases of non-incident overlap in form or implicitly assume the overlap in form to be non-incident. However, there exist interlingual homographs and homophones, which just happen to have the same the orthographic form (e.g., the English *windy* and the Polish *windy*, meaning “elevators/lifts”) or pronunciation (allowing, of course, for phonetic differences), for example, English *sock* and Polish *sok* (“juice”). The meaning in such cases is likely to be unrelated. One point that is of high relevance to the present study is that the orthographic and phonetic similarity between cognates does not necessarily coincide. A pair of words may look strikingly similar in writing, for instance, but not sound the same, for example, English *psyche* and French *psyché* [psif].

Similarity of form is only one aspect of cognate status. The situation is further complicated by the fact that, semantically, a pair of cognates may be in any relationship, from an (almost) complete overlap in meaning, through all degrees of similarity in meaning, to a completely different meaning (for in-depth discussions of the different kinds of similarity between cognates, see Ringbom [2007] and Otwinowska [2015]). The last possibility, in the case of etymologically related words, is the result of various processes of semantic divergence that take place over time. Such pairs of lexical items, similar-looking or sounding but with different meanings, are often referred to as “false friends,” “false cognates,” or “deceptive words.” Although the status of cognate/false cognate is often presented as a dichotomy, it needs to be remembered that this is only a useful simplification. Cognate words in two languages, even when close in meaning, may differ in sometimes very subtle ways, such as in semantic prosody or collocational patterns. Some false cognates, on the other hand, may not be entirely “false”: their deceptive properties may relate only to one of the many meanings of a word. For example, the Polish verb *kolaboracja* shares with the English *collaboration* its disapproving sense, that is, the meaning of “traitorous cooperation with an enemy” (OED). However,

the other sense of the English word (listed as the main meaning by the *OED*), that of “the action of working with someone to produce or create something,” is not shared by the Polish word.¹ This may lead to misunderstanding or misuse in some cases, but it may have a facilitating effect for acquisition in other cases. However, for most purposes, it is useful to distinguish between cognates and false cognates, disregarding borderline cases, and this is the approach taken in this study.

The potentially misleading nature of false cognates has been documented by researchers. False cognates are likely to be wrongly assigned the same meaning(s) in the L2 as in the L1 (Brenders, van Hell, & Dijkstra, 2011; Vidal, 2011). The fact that false cognates create problems in L2 acquisition has been used to support the view that L1/L2 similarity does not necessarily facilitate the acquisition and use of the L2 (Schachter, 1974; see also Granger [1993] for a discussion).

The processing of cognates has been studied in connection with lexical access in bilingual and multilingual speakers, often showing a “cognate advantage” or “cognate facilitation effect.” In isolated word-recognition tasks, cognates and non-cognates elicit different response patterns. For example, cognates are recognized faster in lexical decision tasks (Dijkstra, Grainger, & van Heuven, 1999; Lemhöfer & Dijkstra, 2004; Lemhöfer, Dijkstra, & Michel, 2004). This effect is not limited to just two languages, and the direction is not limited to the L1 affecting the L2; however, a certain level of proficiency must be reached in the second or third language for cognate advantage to occur in the L1 (van Hell & Dijkstra, 2002). Translating cognates has also been shown to take less time than translating non-cognates (de Groot, 1992; Sánchez-Casas, García-Albea, & Davis, 1992).

Substantial evidence of cognate effects is provided by a large body of research involving priming, both unmasked and masked (for unmasked priming: Bowers, Mimouni, & Arguin, 2000; for masked priming: de Groot & Nas, 1991; Gollan, Forster, & Frost, 1997; Sánchez-Casas et al., 1992). On the basis of a series of experiments involving Greek/French cognates (i.e., from languages with different scripts), Voga and Grainger (2007) conclude that two mechanisms combine to create the cognate advantage in priming studies: first, translation primes (both cognate and non-cognate) produce facilitation because of their shared meaning representation, and second, this semantic priming is augmented by a form-priming component in the case of cognate words. Midgley, Holcomb, & Grainger (2011) note that cognates in the L2 benefit from form-meaning associations that pre-exist in the L1, which explains why studies typically find a strong cognate advantage not only in L2 lexical processing but also in L1 lexical processing; the

increased exposure to the same orthographic and/or phonological patterns further adds to this effect.

Studies on bilingual lexical access, such as those mentioned above, are typically carried out with bilinguals and multilinguals, and interpret the significance of the observed patterns of the cognate facilitation effect with respect to theoretical models of bilingual processing. Of equal (if not greater) relevance to this article are studies involving what [Vanhove and Berthele \(2015b\)](#) call “cognate guessing tasks.” In such studies, the participants try to guess the meaning of cognates in a language they do not know. Such studies seek to establish what factors are conducive to the development of receptive multilingualism (see [Vanhove and Berthele \[2015a\]](#) for an overview). [Vanhove and Berthele \(2015a\)](#) point to some item-related determinants of cognate guessing: the extent of the similarity between the cognates; the importance of consonants, which seem to contribute more to the perceived transparency of cognates than vowels; the role of word beginnings, to which participants appear to be more sensitive than to the other parts of lexical items; and the corpus frequency of cognates, with high-frequency words being easier to recognize than low-frequency words.

Researchers generally recognize the fact that formal similarity of a new language with an already known one is helpful for L2 learning ([de Groot & Keijzer, 2000](#); [Lotto & de Groot, 1998](#); [Odlin, 1989](#); [Ringbom, 2007](#)) and that it may help especially with the development of “passive” knowledge of a language ([Ringbom, 2007](#)). It has been suggested that, in L2 learning, cognates can function as “a learner’s first foothold into the new lexicon,” because of their shared form and meaning across languages ([Midgley et al., 2011](#), p. 1634). Indeed, there is research-based evidence that cognates are learned more easily than non-cognates, even though some findings are contradictory ([Otwińska, 2015](#); [Tokowicz, 2015](#)). A recent study found that although the form–meaning connection for cognates seems to be easier to learn, learners may be more reluctant to use cognates in context than non-cognates ([Rogers, Webb, & Nakata, 2015](#)), a finding that the authors of the study tentatively attribute to the possibility that their learners may initially lack confidence when using cognates in context, as they may be aware that the overlap in meaning may not be complete.

An important issue that needs to be mentioned here is that the similarity between words in two languages may not be equally noticeable to all language learners. Indeed, cognate awareness – the ability to recognize the relation between cognates in two languages – has been studied in its own right, as a type of metalinguistic awareness. Despite some contradictions, the body of research on this issue ([Hipfner-Boucher & Chen, 2016](#)) indicates that cognate awareness supports the

transfer of semantic knowledge from L1 to L2 and thus also contributes to improved word reading, vocabulary learning, and reading comprehension. It has been suggested that while the awareness of cognates is not automatic, it can be increased by training (Otwińska-Kasztelanica, 2011). Currently, a growing number of researchers argue in favour of cognate awareness raising as a useful technique to improve vocabulary learning (Otwińska-Kasztelanica, 2009, 2011; White & Horst, 2012).

Language distance: English/French versus English/Polish

As noted above, the extent of cross-linguistic similarity is determined mostly by language typology. English, French, and Polish all belong to different branches of Indo-European (Germanic, Romance, and Slavic, respectively), but even though English and French do not belong to the same group, they share a vast number of lexical similarities due to the imposition of French on speakers of English in the aftermath of the Norman Conquest. The fact that more contemporary English words come from French than from Germanic roots, and that about 58% of English words come from either French or Latin (Williams, 1986), makes English more akin to a Romance language than to a Germanic one. LeBlanc and Séguin (1995) identified 23,160 French/English cognate pairs in dictionaries that totalled some 70,000 entries, suggesting that cognates represent at least 30% of the English vocabulary.

The situation is quite different when English and Polish are compared. Although Polish is a Slavic language, the typological distance from English does not preclude the existence of cognates, due to the borrowings that Polish made from Latin and modern European languages (Otwińska-Kasztelanica, 2009). However, the overlap in vocabulary is not nearly as extensive as the one between English and French. While the exact extent of Polish/English cognate vocabulary is not known, the number of over 3,000 cognate pairs, mentioned by Otwińska-Kasztelanica in several publications on the topic (Otwińska-Kasztelanica, 2010), gives a rough estimate. A dictionary of English/Polish cognates intended for learners of English, by the same author, lists 2,747 cognate pairs, a vast majority of them nouns (Otwińska-Kasztelanica, 2004).

Examination of the Peabody test

Despite a handful of studies on possible racial or ethnic bias (Champion, Hyter, McCabe, & Bland-Stewart, 2003; Webb, Cohen, & Schwanenflugel, 2008), no investigation seems to have been conducted on possible L1 bias in the Peabody Picture Vocabulary Test (henceforth PPVT). Platt (2010) analyzed the first 72 items of the third version of

the PPVT (Dunn & Dunn, 1997), among which seven items showed DIF between pre-schoolers who were speakers of English as an L1 and pre-schoolers who were speakers of English as an L2 and whose first languages included Korean, Urdu, Farsi, and Mandarin. However, those results remain inconclusive due to small sample size and large standard error. In addition, even if Platt's results proved conclusive, they would hardly be generalizable to adult speakers of English as a second language, given the young age at which many of those children started learning English, because they could be classified as cases of bilingual first language acquisition rather than second language acquisition (De Houwer, 2009). In addition, having participants with two different L1 backgrounds would facilitate exploration of L1 bias on the test. This is what the present study aims to do.

Purpose

The purpose of this study is to examine if L1 bias may occur when the PPVT-IV (i.e., the fourth edition of the PPVT) is administered to ESL learners, due to the presence of cognates, using a combination of statistical techniques.

To explore the impact of cognate status on the performance on the PPVT, we selected examinees who speak languages that are markedly different in the number and kinds of cognates they share with English: speakers of French and speakers of Polish. We expect the different number of cognate items to result in bias in favour of the French-speaking participants.

Method

Participants

Our study was conducted with a total of 443 participants, of whom 293 were French-speaking residents of Quebec and 150 were native speakers of Polish, from southern Poland. All were adult university students. Participants were 70% ($n = 310$) female and 30% ($n = 133$) male, with females outnumbering males in both language groups (approx. 2:1 in Quebec and 3:1 in Poland). Given research evidence that the ability to guess the meaning of cognates evolves with age (Vanhove & Berthele, 2015b), it was important that our participants be comparable in terms of age. A *t*-test shows no difference ($t = 1.58$, $p = 0.12$) between the mean age of the Quebecers (20.6) and that of the Poles (20.2).

Our participants were intact groups taking English language courses at the university level. The groups were selected after a careful consideration of the ways in which their level could be verified. The

Table 1: Participants' level of English

Group membership:	Polish-speaking participants (%)		French-speaking participants (%)	
	B1 (<i>n</i> = 54)	B2 (<i>n</i> = 46)	B1 (<i>n</i> = 48)	B2 (<i>n</i> = 52)
Self-evaluated	A2 (8)	B1 (10)	A2 (6)	A2 (4)
level within the	B1 (78)	B2 (76)	B1 (79)	B1 (8)
group:	B2 (14)	C1 (13)	B2 (12)	B2 (74)
			C1 (3)	C1 (14)

majority of the students were assigned to groups on the basis of various placement tests. Such individual test results could not be used for the purposes of the study, as the placement tests were not directly comparable within or across language groups, and their results were in some cases confidential. However, general placement test information in combination with official course descriptions, as well as detailed information from teachers and program coordinators, allowed us to establish which language groups at each institution best corresponded to the B1 and B2 levels on the Common European Framework of Reference for Languages (CEFR; Council of Europe, 2011). The intended composition for both the Polish and the Canadian participants was 50% B1 and 50% B2. However, due to the fluctuating sizes of the intact groups and to absences on the day we tested them, the proportion of B1 and B2 participants was in fact 54/46% for the Polish participants and 48/52% for the Canadian participants (see Table 1).

To ensure equivalence across language groups, the English level of the participants was further verified by means of a self-rating procedure. The participants were presented with a description of the abilities corresponding to each of the six levels of the CEFR and were asked to self-evaluate their competence in English. As expected, a vast majority of the participants rated themselves as either B1 or B2, with only a few participants rating themselves lower (A2) or higher (C1) (see Table 1). The results of the self-rating procedure confirmed our assumption that the participants can be treated as belonging to the targeted competence ranges (B1 and B2), in comparable proportions.

In both countries, the participants had learned English at school, from the elementary level. We also collected data on the participants' knowledge of languages other than English. About half (*n* = 77) of the Polish participants reported some knowledge of German, while in the case of Quebecers the most popular third language was Spanish, mentioned by a quarter (*n* = 80) of the participants. Various other languages were reported to a much lesser extent in both countries. However, the information on the participants' third or other languages was

not taken into consideration, since a vast majority of our participants self-rated their third language at the beginner or lower-intermediate level of proficiency. In view of the research findings cited above, that a certain level of proficiency is required for the source language of a bilingual person's many languages to facilitate the guessing of cognates in another language, it seemed rather unlikely that any such influence would be noticeable. Given the large number of different languages learned as third or fourth ones by our participants (nine in Poland, eight in Quebec), we could eliminate the possibility of a specific third language affecting the result of a whole group in any significant way.

The PPVT-IV

The Peabody Picture Vocabulary Test, fourth edition (PPVT-IV; [Dunn & Dunn, 2007](#)) is a standardized tool widely used to measure receptive vocabulary in American English. The fourth edition, containing 228 items, is deemed suitable by its designers for testing non-native speakers of English. The test-taker is required, for each item, to select from among four colour images the one that reflects the meaning of a word read by the examiner. Designed for individual testing, the test needs to be stopped by the examiner as soon as eight errors have been made within a 12-word set. It is recommended to adjust the starting point according to the age of the examinee when testing native speakers of English, but this recommendation does not apply to non-native speakers, who have to start at the beginning of the test, thus making group testing possible with non-native speakers.

For the purposes of this study, whether a test item is a cognate or not was established by means of ratings by competent bilingual judges, which is a common, widely adopted method of cognate identification ([Otwinowska, 2015](#); [Tokowicz, 2015](#)). There exist measures designed specifically to assess the objective formal similarity of cognates, such as Levenstein Distance (LD); however, since it measures orthographic similarity, and the stimulus words in this study were read out, preference was given to ratings by three judges, who were explicitly instructed to go by the spoken form of the word if in doubt as to its cognate status.

Of the 228 items comprising the Peabody test, 117 are cognates with French, while only 41 are cognates with Polish. However, there is an overlap of 35 items between those two figures – that is, those 35 items are cognates for both language groups. On the other hand, eight test items are false cognates for French, and eight are false cognates for Polish. Only one item is a false cognate for both groups. Considering the figure of 30% of shared cognates between English and French mentioned above, we can observe that words which are cognates with

French are overrepresented on the PPVT-IV (they constitute 51% of all items).

Procedure

Prior to the testing sessions, the teachers who agreed to participate in the study forwarded to their students a description of the abilities corresponding to each of the six levels of the CEFR. Students indicated their self-evaluated level on the answer sheet at the time of testing.

As discussed above, the fact that non-native speakers were tested allowed for group testing. Although the Peabody was administered individually for a handful of our participants, individual testing would have been too time-consuming to reach a sufficient number of participants. Consequently, a vast majority of our participants were tested in groups of about 20. The test slides were projected on a screen in front of the class, and participants were given an answer sheet with 228 slide numbers. Next to each slide number, they were instructed to write down the letter (a, b, c, or d) that corresponded to what they assumed to be the correct meaning of the word they heard, among the four image choices that are provided on each slide. Sufficient time was allowed for all participants to write down an answer for each item. The examiner kept an eye on the class and would not move on to the next slide if a participant was still working on an item. This method allowed answers to be gathered for the whole test (i.e., 228 items), even though several participants would not have reached the end with the procedure recommended for individual testing. The total testing time ranged from 34 to 42 minutes. No fatigue effect was assumed, since participants were adult university students tested for a length of time that they would normally spend on academic tasks requiring concentration.

Analysis

Three analyses were carried out. The first consisted of descriptive statistics and multivariate statistics. The mean scores of Quebecers and Poles were compared for each of the 19 sections of the test. In addition, correspondence analysis was used to investigate how French and Polish speakers could be classified into two distinct groups. The second analysis aimed at examining the impact of cognates and false cognates for each L1 under investigation. We used the mean scores of Poles and Quebecers to see which cognates and false cognates favour one language group over the other. The third analysis consisted of two well-known differential item functioning (DIF) methods not based on Item Response Theory: Mantel-Haenszel (Holland & Thayer, 1988) and logistic regression (Swaminathan & Rogers, 1990). These methods were

selected because they are among the most widely-used DIF methods in language research. The goal was to identify the items that induce bias by favouring one language group over the other one.

Each analysis was performed using R software.² In our data matrix, a correct answer was awarded one point and an incorrect one was awarded zero points. The software then computed the mean for every item. For example, if an item was identified correctly by 60 out of 100 participants, the mean score for that item was 0.6.

Each item on the PPVT test was given one of three possible labels with respect to its cognate status for the speakers of a given L1: (a) not a cognate (i.e., no similarity); (b) cognate (a “true” cognate: similar form, similar meaning); (c) a “false” cognate (deceptive similarity: similar form, different meaning). This yields nine possible combinations of these labels for two L1 groups, which are presented in Table 2, along with the number of instances for each combination.

Of interest to us are only the pairings in which the two labels are different, because if the status of an item is the same for speakers of both L1s (corresponding to shaded squares in Table 2), that item should not favour one language group over another. We therefore looked at items that belong to the categories not shaded in Table 2, that is, the following types of pairings: (a) not a cognate – cognate; (b) not a cognate – false cognate; (c) cognate – false cognate.

We expected significantly higher scores for the group that has an item as a cognate while the other group does not have it as a cognate (type a above). Conversely, we expected significantly lower scores for the group that has an item as a false cognate while the other group does not have it as a cognate (type b). Finally, we expected even greater difference in scores for type c, where the item would strongly favour the speakers of the language for which it is a cognate, while misleading the speakers of the language for which it is a false cognate.

Table 2: Cognate status for the 228 PPVT items

	Polish		
	not a cognate	cognate	false cognate
French			
not a cognate	95	5	3
cognate	78	35	4
false cognate	6	1	1

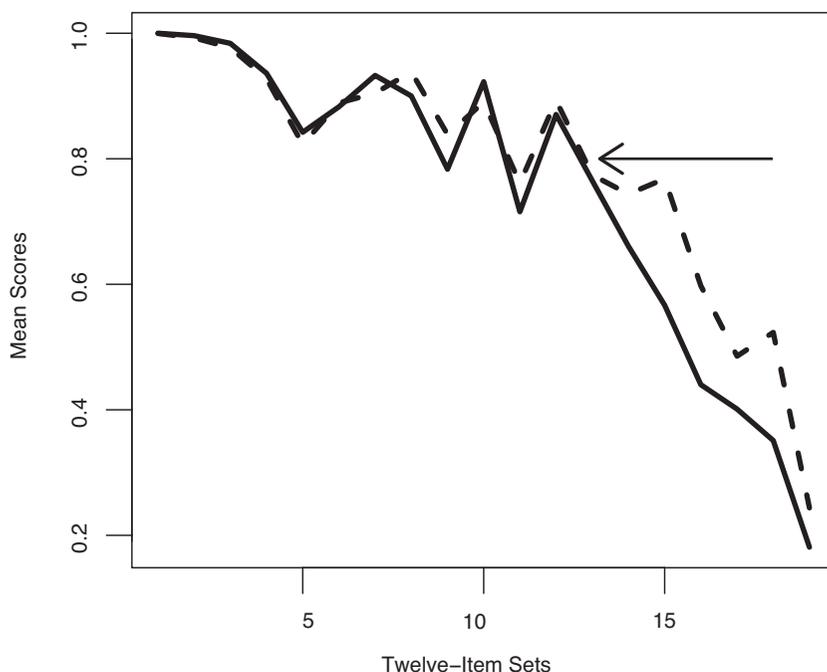


Figure 1: Mean score by Quebecers (dashed line) and Poles (solid line)

Results

Analysis 1: Comparisons of French and Polish speakers

Since the Polish-speaking and French-speaking participants in this study were assumed to be of equal proficiency in English, the first step was to verify if the mean score on the PPVT-IV was equal across groups. Contrary to expectations based on their course levels, a *t*-test showed that the mean score obtained by the French-speaking participants (153.8) was significantly higher than that (137.0) obtained by their Polish counterparts ($t = 8.48, p < 0.01$). This difference might be caused by the nature of the test, giving an advantage to the French-speaking participants in the form of more numerous cognates than for the Polish participants.

Figure 1 shows the mean scores to be equivalent for the first sets of items of the PPVT-IV. However, Figure 1 also shows that native French speakers tend to have higher mean scores on the last sections of the test, starting at set #13 (i.e., item 145 and above). This cut-off point, at which French scores become significantly higher than Polish scores, is indicated by an arrow.

Table 3: Top 10 items that favoured French-speaking examinees

	Item #	French speakers	Polish speakers	Difference (%)
<i>vitreous</i>	221	0.61	0.04	-93
<i>sternum</i>	188	0.79	0.23	-71
<i>submerging</i>	179	0.79	0.27	-66
<i>parallelogram</i>	165	0.94	0.32	-66
<i>convex</i>	213	0.47	0.17	-63
<i>incarcerating</i>	190	0.93	0.35	-62
<i>feline</i>	149	0.95	0.36	-62
<i>timpani</i>	207	0.59	0.23	-62
<i>incandescent</i>	193	0.82	0.33	-60
<i>cornea</i>	173	0.91	0.37	-20

To extract more specific information from the data, we took a closer look at means for every item on the test. When doing so, one must keep in mind that a 0.20 difference is less important when the means are, for example, 0.70 and 0.90 than when they are 0.10 and 0.30. A group with a mean of 0.90 is only slightly stronger than a group with a mean of 0.70, whereas a group with a mean of 0.30 is three times stronger than a group with a mean of 0.10, even though the raw difference is the same in both of these hypothetical cases. Consequently, the differences observed were adjusted as a percentage of the higher of the two scores. As seen in [Table 3](#), for example, the mean score obtained by Poles (0.04) for the first item on the list is 93% lower than the mean score obtained by the stronger group, here the Quebecers (0.61).

In the first column of [Tables 3](#) and [4](#), we italicized the items for which the cognate status was different for the two language groups. As [Table 3](#) shows, all the items that most favoured French speakers were predicted to do so on the basis of their cognate status.

Of the top 10 items that favoured French-speaking participants, nine are clear cognates, while the tenth item (*timpani*) could be guessed by visual and phonetic analogy with the French word for eardrum (*tympan*), which has the same root and shares similarities.

The same comparison identified the items that most favoured the Poles, as represented in [Table 4](#). This time, the italicized items show that four of the top 10 influential items were predicted to favour the Polish speakers, based on their cognate status with Polish (*hydrant*, *sedan*) or to their false cognate status with French (*hydrant*, *talon*, *vest*).

Among the top 10 items that favoured Poles, we find the only two items that are cognates with Polish but not with French (*hydrant*, *sedan*), along with three of the only eight false cognates with French (*hydrant*, *talon*, *vest*).

Table 4: Top 10 items that favoured Polish-speaking examinees

	Item #	Polish speakers	French speakers	Difference (%)
<i>hydrant</i>	115	0.90	0.27	-71
<i>tonsorial</i>	228	0.19	0.06	-67
<i>talon</i>	183	0.18	0.06	-67
<i>tusk</i>	130	0.50	0.18	-65
<i>garment</i>	147	0.77	0.32	-59
<i>vest</i>	59	0.99	0.42	-58
<i>sedan</i>	162	0.68	0.30	-56
<i>upholstery</i>	196	0.13	0.06	-52
<i>hedge</i>	150	0.57	0.30	-47
<i>hovering</i>	153	0.54	0.30	-44

To understand more precisely how Quebecers and Poles provided specific answer patterns, we created a scatterplot of individual factors scores. Plotting the chi-squared data in a two-dimensional space (the rows and columns of our data matrix being the two dimensions) helps visualize the association between items belonging to separate groups. An independence in response patterns is seen in [Figure 2](#). Here, the Quebecers are represented by circles and the Poles by triangles. As we can see, these groups are clearly in opposition: Scores within each language group show more proximity to one another than to scores from the other language group (i.e., the circles and the triangles tend to be located in two distinct areas of the plot). This suggests the existence of two groups of response patterns, corresponding by and large to the two language groups.

Analysis 2: Cognates and false cognates

In our quest for L1-based differential item functioning, we analyzed only the items from sections 9 to 19. The reason for this selection is that for the first eight sections, scores were too high and similar across groups, creating a ceiling effect that prevented the data from showing significant DIF. Section 9 (beginning with item 97) is the first section where mean scores drop below 0.90 for both groups (see [Figure 1](#)). This 90% figure has been advocated as a cut-off point under which performance becomes less stable (cf. [Brown, 1973](#)). Although cognates appear to be evenly distributed throughout the test for each language, they are not assumed to have any impact for the first eight sections, due to the ceiling effect stemming from high item frequency.

It should be noted that when an item is a cognate for both languages, it was not included in our analyses, because we could not expect that item to favour one language group over the other.³

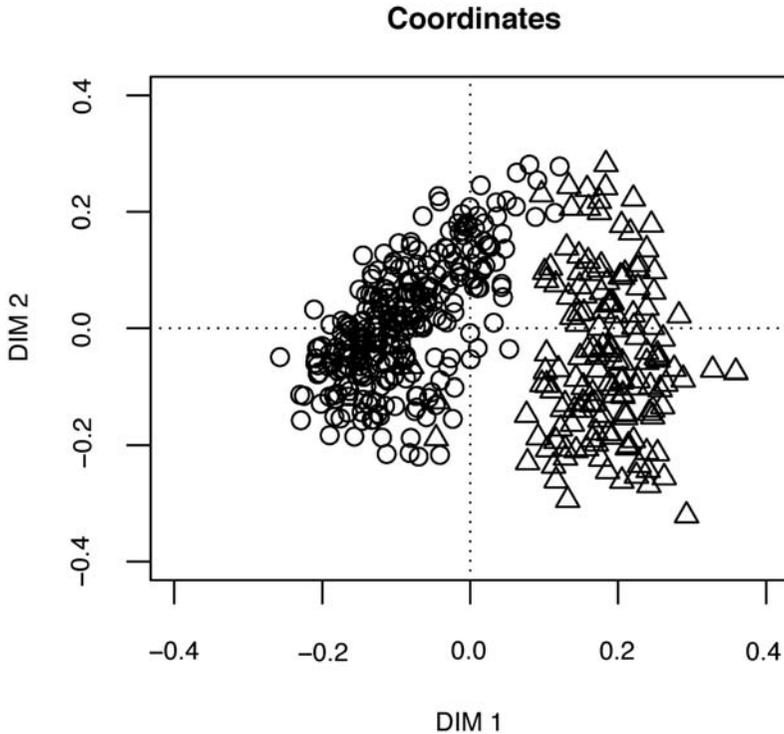


Figure 2: Correspondence analysis – respondents only (Quebec: circles; Poland: triangles)
 Note: DIM 1 = Dimension 1; DIM 2 = Dimension 2

Consequently, after removing the 35 items that are cognates for both language groups, we obtained 77 cognates with French only and five cognates with Polish only (see [Table 2](#)).

In [Figure 3](#), the left panel shows the scores of our respondents for the items that are cognates with French only. The coordinates on both axes represent the mean item scores for each group, and the numbers plotted represent the test items, from 0 to 228. Unsurprisingly, the mean scores of these items were generally higher for the French speakers, being gathered below the diagonal line corresponding to equal scores between the two groups. The right display in [Figure 3](#) shows the scores of our respondents for the Polish cognates. Only five items of the PPVT-IV are cognates with Polish but not with French. Despite that cognate status, two of the five items yielded a higher mean for French-speaking participants (items 54 and 88).

[Figure 4](#) presents the scores for the false cognates for French speakers and Polish speakers. As with the cognates, if an item was a false

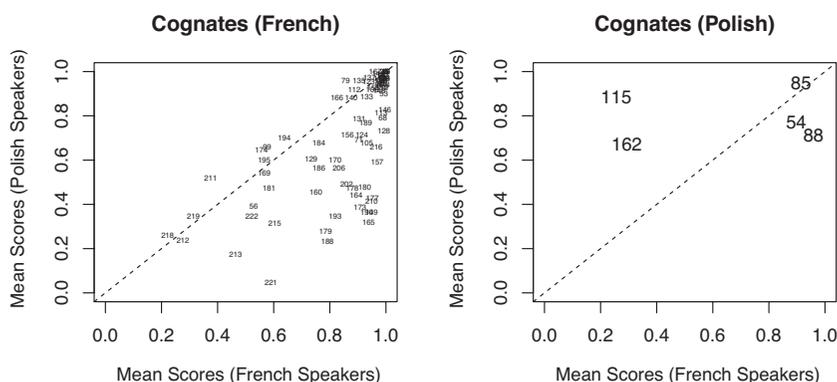


Figure 3: Mean scores for French cognates (left panel) and Polish cognates (right panel)

cognate for both languages, it was not included in the analysis (there was only one such item, 217). In the left panel of this figure, we see that items 115 (*hydrant*) and 59 (*vest*) are clearly in favour of the Poles. Interestingly, item 115 (*hydrant*) is not only a false cognate for speakers of French but also a cognate for the Polish speakers. Likewise, item 161 (*transparent*) is a false cognate with Polish but a cognate with French. In those two cases, therefore, the two types of cross-linguistic effects in a single item may be contributing to a particularly high discrepancy between the two language groups. Both items are farthest from the line in their respective diagrams.

Analysis 3: Differential item functioning

Output analysis for single items in sections 9 to 19 showed a high number of DIF items: 58% of items were detected as DIF using the Mantel-Haenszel method and 63% were detected as DIF using logistic regression (see Appendix). There was considerable agreement between the two methods, such that 55% of the items (73 out of 132) showed DIF by both methods.

To compare the results of the DIF analysis with the earlier analysis of the impact of cognates on the test-takers' performance, we selected those 73 items for which DIF was detected using both methods (Mantel-Haenszel and logistic regression). These items are marked with bold type in Table A1 in the Appendix. We compared them with items for which the cognate status of the target word was different for the two language groups. Those items are shaded in Table A1. Of those 73 items, 59 belong to type a (see the Analysis section); that is, they are cognates for only one of the two languages, with no similarity with the other language. While 55% of all the 132 analyzed items

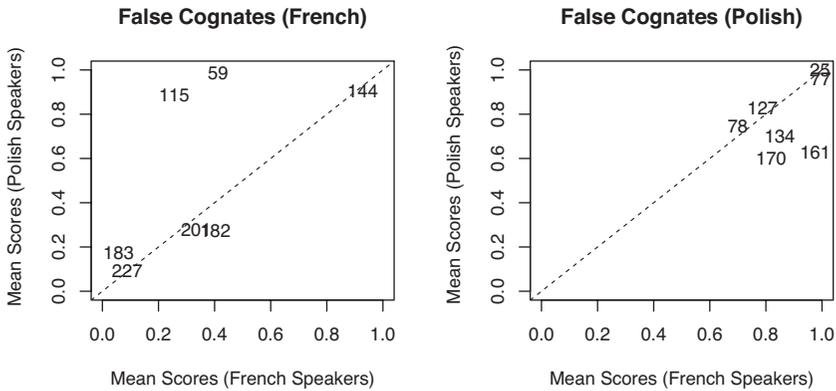


Figure 4: Mean scores for French false cognates (left panel) and Polish false cognates (right panel)

showed DIF, when we analyzed the 59 items that we expected to favour the language group for which they were cognates, this percentage climbed to 66% (39 items out of 59). For type b (false cognates for one language, no similarity with the other language) and type c (false cognates in one language and cognates in the other), the numbers of such items (9 and 5 respectively) are too low to provide any reliable figure.

Discussion and conclusion

This study examined cognate effects in the Peabody test, a test of receptive English vocabulary. The test was administered to speakers of French and to speakers of Polish, two languages that vary in the number and nature of cognates they share with English. We compared the mean scores for both language groups, analyzed the impact of cognates and false cognates, and examined differential item functioning.

Our results show a clear effect of word form: cognates exerted an influence on test performance, thus supporting the bulk of research on cognate effects (de Groot, 2010). Like participants in the studies conducted by Chen and Henning (1985), Sasaki (1991), and Stoeckel and Bennett (2013), our participants whose L1 shared a high number of cognates with the language being tested scored higher on those cognates than did other participants who did not have those same items as cognates.

The fact that 14 of the 20 most influential items we examined (Tables 2 and 3) are cognates in only one language indicates a notable effect of cognates on test performance, which means that a test like the

PPVT-IV is likely to favour examinees who speak a language with which English shares more cognates over examinees whose L1 shows little vocabulary overlap with English.

It is not easy to identify the reason an item will show DIF and advantage one subgroup of examinees over another (Ferne & Rupp, 2007). However, more than 80% of the items that show a difference in status between the French and Polish participants are cognates that English shares with French. This suggests that for test-takers with the same level of English competence, vocabulary-based tests such as the Peabody could tend to overrate speakers of French by allowing them to use L1-reference strategies for guessing the meaning of many unknown items. Although this remains to be quantified, less common English words might possess Latin roots more often than English words of high frequency, thus helping French speakers to guess the meaning of rare English words simply due to their knowledge of a Romance language.

In addition, even though the number of test items that were false cognates in one of the two languages was too low (15) to provide any reliable and generalizable figures, this issue raised an interesting point. We identified items as being false cognates based on the shared similarities in spoken form with French or Polish, as advocated by previous research. As mentioned earlier, false cognates are a hurdle for L2 learners, who will tend to assign to the L2 word the meaning that a similar word has in their native language. Although this phenomenon is well documented and proves true in authentic communication (Dijkstra et al., 2015; Midgley et al., 2011), false cognates may not have the same effect in all testing situations, because a wrong L1-based interpretation may not be offered by the test itself. For example, a French speaker hearing the English sentence “I saw a barb” (item 201) might assume that it refers to a beard (*barbe*, in French). However, when being tested on the PPVT-IV, the same French speaker would not see an image of a beard among the four choices for item 201, so there would be no reason that he or she would be misled by that false cognate into choosing a specific answer that is different from the correct one. In such cases, the examinee would be likely to disregard the similarity in spoken form, resulting in the absence of a negative impact by that false cognate. The same could be said about item 134 (*links*, meaning the rings in a chain): this word could be misinterpreted by Poles in a communicative context as a false cognate, when taken to mean Internet hyperlinks. However, since there is no image choice for that item related to URLs or computers, the similarity in form would likely be ignored here too. Our results seem to support this assumption: Although only eight items were false cognates for only one

language, only three of them showed DIF. Probably by pure coincidence, for one of those two items (*talon*, meaning “heel” in French), one of the alternative choices showed what could be misconstrued as a heel, which misled the French speakers into selecting that wrong answer. For those reasons, when examining test validity, the positive impact of cognates is stronger than the negative impact of false cognates.

It should be noted that this study assumes that words can be classified as either cognate or non-cognates, in a binary fashion. Admittedly, this involves a certain simplification, which we think is justified, but it needs to be acknowledged that at a very fine-grained level of analysis, different degrees of cognateness could be distinguished, as some cognates are identical, some are almost identical, and some bear a similarity that may not be noticeable to everyone (Otwinowska, 2015). In this study, we classified as cognates or false cognates those items that bear a clear resemblance to words in the other language.

What has been said in this article is by no means to be taken as a criticism of the Peabody test. There is no way that cognates could be avoided in a test of this type – its authors would have to take into consideration all of the world’s languages at the same time, which is hardly a feasible task. Rather, in view of the fact that cognates are unavoidable, and a different set of cognates and false cognates will emerge for every pair of languages, we are arguing that results of tests such as the PPVT should be interpreted with caution when it comes to speakers of different mother tongues. Speakers of a language that shares a large number of cognates with the tested language may appear stronger than they actually are. We believe that the results of this study warrant further research into the impact of cognates and false friends on psychometric qualities of language tests such as dimensionality, person-fit, or equating.

Correspondence should be addressed to **Justyna Leśniewska**, Institute of English Philology, Jagiellonian University, Al. Mickiewicza 9A, 31–120 Kraków, Poland; e-mail: justyna.lesniewska@uj.edu.pl.

Notes

- 1 This meaning is listed by some Polish dictionaries, but always as archaic.
- 2 It is possible to obtain the R code by contacting the third author of this article, Sébastien Béland, at sebastien.beland@umontreal.ca.
- 3 The mean for those 35 shared cognates was 0.93 for the Poles and 0.94 for the Quebecers. The difference was not significant ($p = 0.08$).

References

- Bardel, C. (2015). Lexical cross-linguistic influence in third language development. In H. Peuhkert (Ed.), *Transfer effects in multilingual language development* (pp. 111–128). Amsterdam, The Netherlands: John Benjamins. <https://doi.org/10.1075/hsl.4.05bar>
- Bowers, J.S., Mimouni, Z., & Arguin, M. (2000). Orthography plays a critical role in cognate priming: Evidence from French/English and Arabic/French cognates. *Memory & Cognition*, 28(8), 1289–1296. <https://doi.org/10.3758/BF03211829>
- Brenders, P., van Hell, J.G., & Dijkstra, T. (2011). Word recognition in child second language learners: Evidence from cognates and false friends. *Journal of Experimental Child Psychology*, 109(4), 383–396. <https://doi.org/10.1016/j.jecp.2011.03.012>
- Brown, R. (1973). *A first language: The early stages*. Cambridge, MA: Harvard University Press. <https://doi.org/10.4159/harvard.9780674732469>
- Champion, T.B., Hyter, Y.D., McCabe, A., & Bland-Stewart, L.M. (2003). “A matter of vocabulary”: Performances of low-income African American Head Start children on the Peabody Picture Vocabulary Test-III. *Communication Disorders Quarterly*, 24(3), 121–127. <https://doi.org/10.1177/15257401030240030301>
- Chen, Z., & Henning, G. (1985). Linguistic and cultural bias in language proficiency tests. *Language Testing*, 2(2), 155–163. <https://doi.org/10.1177/026553228500200204>
- Cole, N.S., & Zieky, M.J. (2001). The new faces of fairness. *Journal of Educational Measurement*, 38(4), 369–382. <https://doi.org/10.1111/j.1745-3984.2001.tb01132.x>
- Council of Europe. (2011). *Common European framework of reference for language: Learning, teaching, assessment*. Cambridge, England: Cambridge University Press.
- de Groot, A.M.B. (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 Groot, A.M.B. (2010). *Language and cognition in bilinguals and multilinguals: An introduction*. New York, NY: Psychology Press.
- de Groot, A.M.B., & 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 Groot, A.M.B., & Nas, G.L.J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, 30(1), 90–123. [https://doi.org/10.1016/0749-596X\(91\)90012-9](https://doi.org/10.1016/0749-596X(91)90012-9)
- De Houwer, A. (2009). *Bilingual first language acquisition*. Bristol, England: Multilingual Matters.

- Dijkstra, T., Grainger, J., & van Heuven, W.J.B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, 41(4), 496–518. <https://doi.org/10.1006/jmla.1999.2654>
- 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(3), 284–301. <https://doi.org/10.1016/j.jml.2009.12.003>
- Dijkstra, T., Van Hell, J.G., & Brenders, P. (2015). Sentence context effects in bilingual word recognition: Cognate status, sentence language, and semantic constraint. *Bilingualism: Language and Cognition*, 18(4), 597–613. <https://doi.org/10.1017/S1366728914000388>
- Dunn, L.M., & Dunn, D.M. (1997). *Peabody Picture Vocabulary Test* (3rd ed.). Minneapolis, MN: Pearson.
- Dunn, L.M., & Dunn, D.M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). Minneapolis, MN: Pearson.
- Ferne, T., & Rupp, A.A. (2007). A synthesis of 15 years of research on DIF in language testing: Methodological advances, challenges, and recommendations. *Language Assessment Quarterly*, 4(2), 113–148. <https://doi.org/10.1080/15434300701375923>
- Gollan, T.H., Forster, K.I., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew–English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(5), 1122–1139. <https://doi.org/10.1037/0278-7393.23.5.1122>
- Granger, S. (1993). Cognates: An aid or a barrier to successful L2 vocabulary development? *ITL: International Journal of Applied Linguistics*, 99–100, 43–56. <https://doi.org/10.1075/itl.99-100.03gra>
- Hipfner-Boucher, K., & Chen, X. (2016). Cross-language transfer of metalinguistic and cognitive skills in second language learning. In X. Chen, V. Dronjic, & R. Helms-Park (Eds.), *Reading in a second language: Cognitive and psycholinguistic issues* (pp. 99–132). New York, NY; London, England: Routledge.
- Holland, P.W., & Thayer, D.T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H.I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Jarvis, S., & Pavlenko, A. (2007). *Crosslinguistic influence in language and cognition*. New York, NY: Routledge.
- LeBlanc, R., & Séguin, H. (1995). Les congénères homographes et parographes anglais-français. In R. Courchêne, S. Burger, C. Cornaire, R. LeBlanc, T.S. Paribakht, & H. Séguin (Eds.), *Twenty-five years of second language teaching at the University of Ottawa* (pp. 69–91). Ottawa, ON: University of Ottawa Press.
- Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition*, 32(4), 533–550. <https://doi.org/10.3758/BF03195845>

- Lemhöfer, K., Dijkstra, T., & Michel, M. (2004). Three languages, one ECHO: Cognate effects in trilingual word recognition. *Language and Cognitive Processes*, 19(5), 585–611. <https://doi.org/10.1080/01690960444000007>
- Lotto, L., & de Groot, A.M.B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, 48(1), 31–69. <https://doi.org/10.1111/1467-9922.00032>
- Matthews, P. (2014). Cognate. In *The concise Oxford dictionary of linguistics*. Retrieved from <http://www.oxfordreference.com/view/10.1093/acref/9780199675128.001.0001/acref-9780199675128-e-554>
- Midgley, K.J., Holcomb, P.J., & Grainger, J. (2011). Effects of cognate status on word comprehension in second language learners: An ERP investigation. *Journal of Cognitive Neuroscience*, 23(7), 1634–1647. <https://doi.org/10.1162/jocn.2010.21463>
- Odlin, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge, England: Cambridge University Press. <https://doi.org/10.1017/CBO9781139524537>
- Ortega, L. (2009). *Understanding second language acquisition*. New York, NY: Routledge.
- Otwinowska, A. (2015). *Cognate vocabulary in language acquisition and use: Attitudes, awareness, activation*. Clevedon, England: Multilingual Matters.
- Otwinowska-Kasztelanica, A. (2004). *Słowa, które znasz. Praktyczny słownik angielsko-polski z ćwiczeniami*. Warsaw, Poland: Wilga.
- Otwinowska-Kasztelanica, A. (2009). Raising awareness of cognate vocabulary as a strategy in teaching English to Polish adults. *Innovation in Language Learning and Teaching*, 3(2), 131–147. <https://doi.org/10.1080/17501220802283186>
- Otwinowska-Kasztelanica, A. (2010). Language awareness in using cognate vocabulary: The case of Polish advanced students of English in the light of the theory of affordances. In J. Arabski & A. Wojtaszek (Eds.), *Second language acquisition studies in neurolinguistic and psycholinguistic perspective* (pp. 175–190). Bristol, England: Multilingual Matters.
- Otwinowska-Kasztelanica, A. (2011). Awareness of cognate vocabulary and vocabulary learning strategies of Polish multilingual and bilingual advanced learners of English. In J. Arabski & A. Wojtaszek (Eds.), *Individual learner differences in SLA* (pp. 110–126). Bristol, England: Multilingual Matters.
- Platt, A.E. (2010). *Examining the appropriateness of the PPVT-III as a measure of vocabulary development in linguistically diverse kindergarten children* (Unpublished doctoral dissertation). University of Toronto.
- Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Clevedon, England: Multilingual Matters.
- 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>
- Sánchez-Casas, R.M., García-Albea, J.E., & Davis, C.W. (1992). Bilingual lexical processing: Exploring the cognate/non-cognate distinction. *European Journal of*

- Cognitive Psychology*, 4(4), 293–310. <https://doi.org/10.1080/09541449208406189>
- Sasaki, M. (1991). A comparison of two methods for detecting differential item functioning in an ESL placement test. *Language Testing*, 8(2), 95–111. <https://doi.org/10.1177/026553229100800201>
- Schachter, J. (1974). An error in error analysis. *Language Learning*, 24(2), 205–214. <https://doi.org/10.1111/j.1467-1770.1974.tb00502.x>
- Stoeckel, T., & Bennett, P. (2013). Sources of differential item functioning between Korean and Japanese examinees on a second-language vocabulary test. *Vocabulary Learning and Instruction*, 2(1), 47–55.
- Swaminathan, H., & Rogers, H.J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361–370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Tokowicz, N. (2015). *Lexical processing and second language acquisition*. New York, NY; London, England: Routledge.
- van Hell, J.G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, 9(4), 780–789. <https://doi.org/10.3758/BF03196335>
- Vanhove, J., & Berthele, R. (2015a). Item-related determinants of cognate guessing in multilinguals. In G. De Angelis, U. Jessner, & M. Kresić (Eds.), *Crosslinguistic influence and crosslinguistic interaction in multilingual language learning* (pp. 95–118). London, England: Bloomsbury.
- Vanhove, J., & Berthele, R. (2015b). The lifespan development of cognate guessing skills in an unknown related language. *International Review of Applied Linguistics in Language Teaching*, 53(1), 1–38. <https://doi.org/10.1515/iral-2015-0001>
- VanPatten, W., & Williams, J. (2008). *Theories in second language acquisition: An introduction*. New York, NY: Routledge.
- Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, 61(1), 219–258. <https://doi.org/10.1111/j.1467-9922.2010.00593.x>
- Voga, M., & Grainger, J. (2007). Cognate status and cross-script translation priming. *Memory & Cognition*, 35(5), 938–952. <https://doi.org/10.3758/BF03193467>
- Webb, M.L., Cohen, A.S., & Schwanenflugel, P.J. (2008). Latent class analysis of differential item functioning on the Peabody Picture Vocabulary Test-III. *Educational and Psychological Measurement*, 68(2), 335–351. <https://doi.org/10.1177/0013164407308474>
- 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>
- Whong, M. (2006). What transfers? In S. Unsworth, T. Parodi, A. Sorace, & M. Young-Scholten (Eds.), *Paths of development in L1 and L2 acquisition: In honor*

- of Bonnie D. Schwartz (pp. 187–199). Amsterdam, The Netherlands: John Benjamins. <https://doi.org/10.1075/lald.39.09who>
- Williams, J.M. (1986). *Origins of the English language: A social and linguistic history*. New York, NY: Free Press.
- Yudes, C., Macizo, P., & Bajo, T. (2010). Cognate effects in bilingual language comprehension tasks. *Neuroreport*, 21(7), 507–512. <https://doi.org/10.1097/WNR.0b013e328338b9e1>

Appendix

Table A1: DIF results

Item	MH	LR	Word	FC	FFC	PC	PFC	Item	MH	LR	Word	FC	FFC	PC	PFC
144		X	dilapidated				X	187		X	depleted				
145	X	X	pedestrian					188	X	X	sternum	X			
146	X	X	interior	X				189			maritime	X			
147	X	X	garment					190	X	X	incarcerating	X			
148			departing	X				191			dejected				
149	X	X	feline	X				192			quintet	X		X	
150	X	X	hedge					193	X	X	incandescent	X			
151	X	X	citrus	X		X		194	X	X	confiding	X			
152			florist	X		X		195	X	X	mercantile	X			
153	X	X	hovering					196	X	X	upholstery				
154	X	X	aquatic	X				197	X	X	filtration	X		X	
155			reprimanding	X				198			replenishing				
156			carpenter	X				199		X	trajectory	X		X	
157	X	X	primate	X				200	X	X	perusing				
158	X	X	glider					201			barb			X	
159	X	X	weary					202	X	X	converging	X			
160	X	X	hatchet	X				203	X	X	honing				
161	X	X	transparent	X			X	204		X	angler				
162	X	X	sedan			X		205			wildebeest				
163			constrained					206	X	X	coniferous	X			
164	X	X	valve	X				207	X	X	timpani				
165	X	X	parallelogram	X				208			pilfering				
166	X	X	pillar	X				209			pestle				
167	X	X	consuming	X				210	X	X	reposing	X			
168	X	X	currency					211	X	X	cupola	X			
169	X	X	hazardous	X				212			derrick	X			
170			pentagon	X		X		213	X	X	convex	X			
171			appliance					214		X	embossed				
172	X	X	poultry					215	X	X	torrent	X			
173	X	X	cornea	X				216	X	X	dromedary	X		X	
174	X	X	peninsula	X				217			legume		X		X
175	X	X	porcelain	X		X		218	X	X	cairn	X			
176	X	X	detonation	X		X		219	X	X	arable	X			
177	X	X	cerebral	X				220	X		supine				
178	X	X	perpendicular	X				221	X	X	vitreous	X			
179	X	X	submerging	X				222			lugubrious	X			
180	X	X	syringe	X				223			caster				
181			lever	X				224			Terpsichorean				
182			apparel		X			225			cenotaph	X			
183	X	X	talon		X			226			calyx				
184			cultivating	X				227			osculating		X		
185			wedge					228	X		tonorial				
186	X		ascending	X											

Note: Bold type indicates the items for which DIF was detected using both methods (Mantel-Haenszel and logistic regression). Shading marks the items for which the cognate status of the target word was different for the two language groups. MH: DIF with Mantel-Haenszel; LR: DIF with logistic regression; FC: French cognates; FFC: French false cognate; PC: Polish cognates; PFC: Polish false cognate.