

# A Corpus Analysis of English-French Neighbour Cognates in L1 French EFL Learner Writing

Xinjingyi Li

Faculty of English Language and Culture, Guangdong University of Foreign Studies

Baiyun Dadao Bei #2, Guangzhou, Guangdong, China, 510420

E-mail: [lysea56@163.com](mailto:lysea56@163.com)

## Abstract

Leveraging cognates has been considered an effective way to teach vocabulary in two languages that are historically and linguistically related, as the similarity in form can help learners with vocabulary retention. However, neighbour cognates, denoting cognates with identical meanings but subtle differences in spelling, may cause negative transfer, as the deceptive similarities in spelling might, at times, impact the accuracy of students' written output. This study aims to shed light on patterns of incorrect spelling forms so as to help learners avoid negative transfer caused by neighbour cognates. Based on a relatively large amount of data from the Longman Learners Corpus (LLC), this study examined and identified patterns of negative transfer resulting from neighbour cognates across all proficiency levels. Following a two-step screening process of 50 randomly chosen neighbour cognates, 26 were included for the final analysis. Four types of negative transfer were identified, including substitution, blends, deviation, and mixed, among which, Type 1 and Type 2 errors have been documented in Odlin (1989) and Ringbom (1987). However, deviation and mixed errors seem not to be well noticed. Besides, Type I error, which involves straightforward substitutions, was found to be the most common in learners' written production, accounting for 77 per cent of the total errors. Therefore, intentional teaching of neighbour cognates, and designing teaching and learning materials as well as vocabulary tests should be encouraged so as to help students notice and avoid these types of errors.

**Keywords:** English-French cognates, lexical transfer, learners' written errors, learner corpus

**DOI:** 10.7176/JEP/15-8-01

**Publication date:** July 30th 2024

## 1. Introduction

Cognates are commonly defined as words in different languages sharing the same or similar phonetic and/or orthographic features and also having identical meanings (Schmitt, 1997, p. 209; Whitley, 2002, as cited in Otwinowska, 2016). Richards and Schmidt (2002) indicate that the occurrence of cognates commonly suggests that these languages are geographically and historically related. For example, English and French, French and Spanish, and so forth.

According to Winford (2003), 75% of modern English words are borrowed from other languages, with the most frequently used 10,000 words being of Romance origin, in which 45% are found to be of French origin (Van Gelderen, 2006). For example, class-classe, legal-légal, government-gouvernement, etc.

Many studies have shown that the existence of cognates can provide a shortcut for language learning, as the similarities in word form and meaning can largely facilitate vocabulary learning (e.g., Cop et al., 2017; D'Angelo et al., 2017; Hipfner-Boucher et al., 2016; Proctor & Mo, 2009; Rabinovich et al., 2018; Van Assche et al., 2013). For example, in D'Angelo et al. (2017), 75 children around 6 years old learning French as either a second (L2) or third language (L3) with various first language (L1) backgrounds, were asked to identify 20 English-French cognates (e.g., dragon-dragon, park-parc) among 40 test items after being taught the concept of cognates. The findings support that even at a very young age, children could be trained to identify cognates and derive vocabulary learning benefits, aligning with another study (Hipfner-Boucher et al., 2016), which also confirms this facilitative role.

Additionally, comparable studies also substantiate the promoting impact of cognates in children or adults possessing varied language proficiency levels, who learn other linguistically related languages, such as English and Spanish (HancinBhatt & Nagy, 1994), Frisian and Dutch (Bosma et al., 2019), Dutch and English (e.g.,

Dijkstra et al., 2010), Polish and English (Otwinowska & Szewczyk, 2019), and so on.

The aforementioned research strongly implies that using cognates is important, and cultivating cognate awareness holds promise in teaching and learning vocabulary, irrespective of diverse language backgrounds and proficiency levels.

Nonetheless, as mentioned at the beginning, not all the cognates in two languages are entirely identical. A specific type of cognates having the same meanings but certain different spellings is referred to as neighbour cognates (of note, these are not false cognates, indicating words exhibiting similarities in pronunciation and/or orthography but differences in meanings). The subtle differences in form of these neighbour cognates can occasionally lead to misspellings, impacting the accuracy of the vocabulary output. Therefore, when the similarity in form of L1 hinders the vocabulary learning of L2, the negative transfer naturally occurs.

However, to the best of my knowledge, most scholars focus on negative transfer arising from false cognates, while largely neglecting the negative transfer resulting from neighbour cognates. Although the number of relevant studies is still relatively small, interest in neighbour cognates has increased, reflecting a growing recognition of the importance of this issue.

In an early study, Albert and Obler (1978) proposed that when words exhibit greater similarities and minimal differences in form, they are at a higher risk of causing confusion, as the increased difficulty in distinguishing them amplifies the likelihood of errors. This standpoint is supported by a series of studies (Odlin, 1989; Ringbom, 1987, 2001). For example, some scholars have specifically pointed out that deviant letter(s) within words (e.g., *government-gouvernement*) often results in a smaller cognate facilitation effect compared to deviations in the final letters of words (e.g., *class-classe*) (e.g., Font, 2001). Meanwhile, the facilitating effects of these two types of neighbour cognates are smaller than those of completely identical cognates (see Comesaña et al., 2018; Font, 2001). In other words, it can be inferred that neighbour cognates, which are more prone to spelling errors, are likely to be more challenging to learn than completely identical cognates.

In summary, making full use of cognates can significantly promote vocabulary learning in two linguistically related languages. However, many scholars might overlook the potential negative transfer caused by neighbour cognates. Moreover, the studies mentioned above have drawn conclusions through relatively limited sample sizes via traditional methods like writing or vocabulary tests, and only a few scholars have examined the facilitative role of identical cognates through corpora (e.g., Lubliner & Hiebert, 2011; Marín & Fernández, 2015). Therefore, by examining a large number of written texts from LLC, this paper aims to demonstrate that negative transfer caused by neighbour cognates is a noteworthy issue.

## 2. Methodology

### 2.1 Resources

#### 2.1.1 Longman Learners Corpus

The Longman Learners Corpus serves as a huge database containing 33,702 writing texts of English learners from varied L1 backgrounds (e.g., Arabic, Chinese, French, etc.) and language proficiency levels (e.g., elementary, intermediate, advanced, etc.). In this paper, I tend to use this corpus in CQPweb (Hardie, 2012) to examine misspellings of English learners with L1 French at all proficiency levels due to neighbour cognates, indicating that the total number of texts considered would be 1051.

#### 2.1.2 Neighbour Cognates Selection Reference: *French-English Comparison—Rapid Memorization of University French Level I-IV Vocabulary* (Zhou, 2003)

This book is a vocabulary memory textbook designed based on English-French comparison. The author compiled 2,887 French words required by the university level syllabus that are related to English words in form and/or meaning. Within the scope of this study, two sets of words are identified neighbour cognates, including regularly patterned synonyms with similar spellings (Category 1, see Table 1) and irregularly patterned synonyms with similar spellings (Category 2<sup>1</sup>), such as *government-gouvernement*, *merchant-marchand*<sup>2</sup>).

<sup>1</sup> Category 1 and Category 2 are two categories presented in the book.

<sup>2</sup> The presentation format of cognate pairs in this article, whether in the main text, tables, or appendices, consistently features English words first, followed by their corresponding French counterparts.

Table 1. Category 1: Regularly Patterned Synonyms with Similar Spellings

Category 1	Examples	Total Count <sup>3</sup>
English words (E) =French word (F)-accent	acceleration-accélération age-âge	112
E=F-e	origin-origine pilot-pilote	87
E=F+e	climate-climat salute-salut	22
E=F-r	dispute-disputer dispose-disposer	74
E=F-er	absorb-absorber ruin-ruiner	66
Total	--	361

It is assumed that the first type in Category 1 would not cause negative transfer, as English is an accent-free language. Of note, although words in Category 1 may have a lower error probability than those in Category 2 due to their regularity, I still included the rest of the four types in Category 1 in the sampling to ensure more comprehensive results. In summary, within the scope of this paper, this paper included the latter four types in Category 1 (a total of 249) and Category 2 (a total of 917 words), totalling 1,166 words.

## 2.2 Procedure

Based on the selection reference (Zhou, 2003) and to guarantee the data is manageable, 50 neighbour cognates were randomly selected from 1166 words. Since 24 pairs were examined not meeting the inclusion criteria, the remaining 26 neighbour cognates were finally presented and analysed in the *result and discussion* part. The detailed selection and screening steps are as follows.

Firstly, since I randomly selected 50 neighbour cognates from a pool of 1166 words, the sampling ratio is 50/1166. I sampled words under different ratios for each type (see Table 2), and the final 50 words for the query are detailed in Appendix 1.

Table 2. Sampling of Each Type

Classification	Sampling Formula	Total Count
F-e=E	87*50/1166	4
F+e=E	22*50/1166	1
F-r=E	74*50/1166	3
F-er=E	66*50/1166	3
F≈E	917*50/1166	39

Secondly, since my aim is to analyse the patterns of spelling errors caused by neighbour cognates, the analysis is based on the premise that the sampled words have a non-zero frequency of occurrence in the corpus and have forms with spelling errors. Therefore, I conducted a two-step screening of the randomly sampled data, ultimately identifying 26 neighbour cognates to be included in the analysis.

In the initial screening, *French* was selected as the *native language category* in *restricted query* to examine whether the correct forms of the randomly selected words had been utilized in the corpus. For example, among the 50 words, neighbour cognates such as *balloon-ballon*, *genu-genou*, *formula-formule*, and so forth were not

<sup>3</sup> The quantity of each type is not calculated manually but indicated in the book.

found in corpus searches, and thus, they were excluded from the analysis.

In the second screening step, wildcards and *frequency breakdown* were used to query the lemmas of neighbour cognates. It is posited that the source of negative transfer lies in the differing spellings within neighbour cognates. For example, in the cognate pair *abundance-abondance*, there is a difference between the letters *u* and *o*, so the errors might likely occur in this position. Therefore, {ab\*ndance} was used in searching. As the results indicated only the correct spelling form in the corpus (see Figure 1), this cognate pair was not included in the subsequent analysis. Similarly, pairs without misspelled forms were all excluded from the analysis. At this point, a one-step random sampling and a two-step screening were completed. Finally, 26 neighbour cognates were included for analysis, as shown in Table 3.

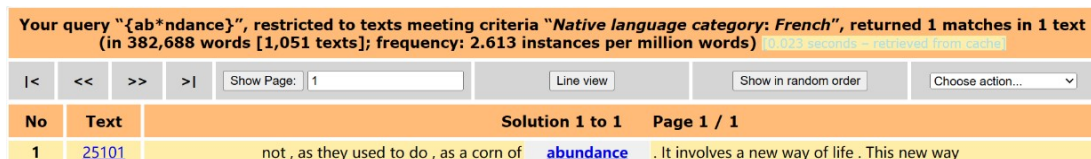


Figure 1. Query Result of {ab\*ndance}

Table 3. 26 Neighbour Cognates Conforming to the Inclusion Criteria

Neighbour Cognates	Neighbour Cognates
address-adresse	board-bord
change-changement	class-classe
combatant-combattant	comfortable-confortable
company-compagnie	complex-complexe
consumption-consommation	department-département
development-développement	domain-domaine
efficacy-efficacy	enemy-ennemi
example-exemple	group-groupe
independent-indépendant	literature-littérature
method-méthode	passport-passeport
personal-personnel	process-processus
republic-république	sympathetic-sympathique
tissu-tissue	treatment-traitement

Moreover, to largely ensure the results only present the lemma of the target word pairs and encompass all possible misspelled forms, I take another example to further illustrate the query process. For instance, in searching for *address-adresse*, as there are two places with spelling differences, it is speculated that potential misspelled English forms might include *adresse*, *adresse*, etc. To largely ensure irrelevant words like *administration* and *advertising* would not occur in the query results, I did not search {ad\*s\*}, {ad\*ss\*} (see Figure 2), etc., but instead used {ad\*res\*} (see Figure 3). The specific query item for each word in this study is detailed in Appendix 2. After completing the searches, the *frequency breakdown* function was used for a more intuitive view of potential misspelled forms, as shown in Figure 3.

Query "{ad\*s\*}", restricted to texts meeting criteria "Native language category: French", returned 256 matches in 133 different texts (in 382,688 words [1,051 texts]; frequency: 668.952 instances per million words)

Showing frequency breakdown of words in this query, at the query node; there are 48 different types and 256 tokens at this concordance position.

Breakdown position: Node | Show: Frequency breakdown of words only | Download frequency breakdown table (for words)

No.	Query result	No. of occurrences	Percent
1	administration	55	21.48%
2	advertising	51	19.92%
3	address	28	10.94%
4	Adjustment	11	4.30%
5	advise	9	3.52%
6	adolescence	8	3.13%
7	advertisement	8	3.13%
8	adress	7	2.73%
9	advertisements	7	2.73%
10	advised	6	2.34%
11	Adams	5	1.95%
12	advertise	4	1.56%

Figure 2. Query Result of {ad\*s\*}

Query "{ad\*res\*}", restricted to texts meeting criteria "Native language category: French", returned 42 matches in 39 different texts (in 382,688 words [1,051 texts]; frequency: 109.750 instances per million words)

Showing frequency breakdown of words in this query, at the query node; there are 6 different types and 42 tokens at this concordance position.

Breakdown position: Node | Show: Frequency breakdown of words only | Download frequency breakdown table (for words)

No.	Query result	No. of occurrences	Percent
1	address	28	66.67%
2	adres	7	16.67%
3	addresses	3	7.14%
4	adreses	2	4.76%
5	addressing	1	2.38%
6	adressed	1	2.38%

Figure 3. Query Result of {ad\*res\*}

It is also crucial to highlight that misspelled words were further checked through concordance lines. For example, in the search results of {ad\*res\*}, *addresses* occurs three times. However, this word form might simply be the plural form or third-person singular form of *address* (see Figure 4), which might not be able to indicate that students misspelled the English word *address* as *adresse* and added an *s*. Similarly, the concordance lines for all search results of the 26 words were scrutinized to ensure that the results discussed in the next section are indeed words with misspellings caused by similarities between neighbour cognates.

Your query "{ad\*res\*}", restricted to texts meeting criteria "Native language category: French", returned 42 matches in 39 different texts (in 382,688 words [1,051 texts]; frequency: 109.750 instances per million words), reduced to results where query node matches word: *addresses* (3 hits)

Show Page: 1 | Line view | Show In random order | Choose action...

No	Text
1	00549 in English . I asked the British Council to give me some <a href="#">addresses</a> in London , and this is how I came to know your
2	01298 'm sending you my salary record , together with the names and <a href="#">addresses</a> of referees whom you may consult for evidence of my experience and
3	13310 doing his political speeches either in Greek or in Spanish whenever he <a href="#">addresses</a> a minority communittee . Moreover , in Indiana , while Jesse Jackson

Figure 4. Concordance Lines of Addresses

### 3. Results and Discussion

The ways of using wildcards for the 26 words and all the error variations are recorded, part of which is shown in Table 4 (see Appendix 2 for details).

Table 4. Query Details (partial)

Cognates Pair	Query	Variation	Frequency
address-adresse	{ad*res*}	address	7
		addresses	2
		addressed	1
board-bord	{bo*rd}	bord	1
combatant-combattant	{comba*ant}	combattants	1
change-changement	{change*/n}	changement	3

By examining the possible spelling errors for the 26 neighbour cognates, four main types of negative transfer caused by neighbour cognates were identified: *substitution*, *blends/hybrids*, *deviation*, and *mixed*. The first two types have been pointed out by some scholars (e.g., Odlin, 1989; Ringbom, 1987), while the third type was originally concluded according to observation. Besides, *mixed* indicates that in some cases, it is uncertain what the specific type of transfer error would be (see Table 5).

Table 5. Four Types of Negative Transfer

Negative Transfer Type	Examples of Misspelling Forms
Type 1: <i>substitution</i>	bord, adresse, changement
Type 2: <i>blends/hybrids</i>	consumation, efficacicy, ennemies (ennemy <sup>4</sup> )
Type 3: <i>deviation</i>	adress, developpment, personnal
Type 4: <i>mixed</i>	adresses (adress/adresse), addressed (adress/adresse),

#### 3.1 Type I: Substitution

Firstly, errors due to *substitution* (Odlin, 1989) are identified (see Table 6). According to Odlin (1989), *substitution* refers to directly copying the L1 lexical form to produce the target vocabulary. This error might occur because learners have not mastered the target vocabulary proficiently. Therefore, during recall, they first search for the corresponding word in L1 and then directly use the L1 word in L2 production. For example, learners directly used the French form *bord* to replace the English word *board*.

Table 6. Examples of Substitution

Cognates Pair	Variations
board-bord	bord
combatant-combattant	combattants
change-changement	changement
class-classe	classe

#### 3.2 Type II: Blends/Hybrids

The second type can be classified as what Ringbom (1987) refers to as *blends* or *hybrids*, indicating a non-existing L2 word is coined by attaching morphemes in the target language to L1 words, leading to incorrect

<sup>4</sup> The words inside the parentheses are the possible base forms of words without agreement.

lexical forms consisting of morphemes from different languages. In this paper, it means students attached English (L2) morphemes to French (L1) words. For example, for the blend *efficacity*, learners might first think of the French word *efficacité*, then attach the nominal bound morpheme *-y* in English and coin the word *efficacity*, which does not exist in English. Other examples found in the corpus are listed below (see Table 7).

Table 7. Examples of Blends/Hybrids

Cognates Pair	Variations
consumption-consommation	consumation
efficacy-efficacité	efficacity
enemy-ennemi	enemies (ennemy)

### 3.3 Type III: Deviation

In addition to the two types of errors mentioned above, some potential errors that do not seem to fit into existing categories were also identified, which were referred to as *deviation*, where the production of the target word is influenced by L1, resulting in a non-existing target word with the addition or omission of a letter(s) (see Table 8). For instance, in the case of *development-développement*, the incorrect form *developpment* is identified. It is possible that learners were impacted by the L1 word *développement*, and the differences in certain letters caused some confusions, so they produced word like *developpment*, which deviates from the correct English word *development* with a letter *p*. It is important to note that this type of negative transfer is distinct from the “totally or partially deceptive cognates” proposed by Ringbom (2001), which are also due to orthographic similarities but result in existing target words.

Table 8. Examples of Deviation

Cognates Pair	Variations
address-adresse	adress
development-développement	developpment
personal-personnel	personnal

### 3.4 Type IV: Mixed

The *mixed* type involves that it is unclear which of the three types of negative transfer mentioned above actually caused the error, which is usually caused by agreement (e.g., plural forms, third-person singular, etc.), as sometimes it is hard to identify the original word (see Table 9). For instance, in the case of *address-adresse*, when the observed error form is *adresses*, it is hard to determine whether students intended to write *adress* (Type 3: *deviation*) or *adresse* (Type 1: *substitution*). To further clarify, we might need to ask the students who produced the incorrect form. However, it is usually very hard to trace the source of data for further verification, which is usually a limitation or drawback of using learner corpora to examine learners’ output.

Table 9. Examples of Mixed

Cognates Pair	Variations	Error Type
address-adresse	adresses (address/adresse)	<i>Substitution</i> or <i>deviation</i>
	adressed (address/adresse)	<i>Substitution</i> or <i>deviation</i>
sympathetic-sympathique	sympathic	<i>Substitution+blends/hybrids</i>

### 3.5 Summary

Among these four types of negative transfer, 77% of them belong to Type I error, 9% belong to Type II error, 11% belong to Type III error, and 3% belong to Type IV error (see Table 10).

Table 10. Descriptive Statistics: Fluency, Percentage

Error Type	Frequency	Percentage
Type I	77	77%
Type II	9	9%
Type III	11	11%
Type IV	3	3%
Total	100	100%

From the above data, it can be generally concluded that Type I error is the most frequent error. Therefore, teachers should spend more time intentionally instructing neighbour cognates causing Type I error. Besides, it is also important to design learning and teaching materials, as well as vocabulary tests targeting various types of neighbour cognates, which can help students notice and avoid these types of errors.

## 5. Conclusion

Research on cognates has shed light on the significance and efficacy of incorporating cognates in both language learning and teaching. However, despite the fact that leveraging the similarity of cognates can largely promote positive transfer, subtle differences in the spelling of neighbour cognates may result in negative transfer. Based on a large amount of written data in the LLC, this study explored and analysed the patterns of negative transfer caused by neighbour cognates. After a two-step screening of 50 randomly selected neighbour cognates, 26 cognates were finally included in the analysis. Four types of negative transfer were identified, two of which are *substitution* and *blends/hybrids*, as mentioned by Odlin (1989) and Ringbom (1987), respectively. The third type of negative transfer, to the best of my knowledge, having not been generalized yet, is categorized as *deviation*. The last one is *mixed*, denoting that it is not certain which type of negative transfer actually caused the error. In addition, among the four error types, it appears that Type I error occurs most often.

## 6. Limitations and Implications

Due to the length limitation, questions like at which language proficiency level errors due to neighbour cognate are more likely to occur, and whether a certain type of error is more likely to occur at a particular proficiency level have not been discussed. Therefore, in subsequent studies, we can further explore the effects of neighbour cognates on students at different proficiency levels, design teaching materials tailored to different types of negative transfer and proficiency levels, and help students better avoid negative transfer due to neighbour cognates.

## References

- Albert, M. L., & Obler, L. K. (1978). *The bilingual brain: Neuropsychological and neurolinguistic aspects of bilingualism*. Academic Press.
- Bosma, Blom, E., Hoekstra, E., & Versloot, A. (2019). A longitudinal study on the gradual cognate facilitation effect in bilingual children's Frisian receptive vocabulary. *International Journal of Bilingual Education and Bilingualism*, 22(4), 371–385. <https://doi.org/10.1080/13670050.2016.1254152>
- Comesaña, M., Coelho, R., Oliveira, H., & Paula Soares, A. (2018). How letter order is encoded in bilingual reading? The role of deviant-letter position in cognate word recognition. *Speech, Language and Hearing*, 21(2), 90–93. <https://doi.org/10.1080/2050571X.2017.1369049>
- 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*, 20(4), 747–769. <https://doi.org/10.1017/S1366728916000213>
- D'Angelo, Hipfner-Boucher, K., & Chen, X. (2017). Predicting growth in English and French vocabulary: The facilitating effects of morphological and cognate awareness. *Developmental Psychology*, 53(7), 1242–1255. <https://doi.org/10.1037/dev0000326>



- Denison, D., & Hogg, R. (2006) Overview. In R. Hogg and D. Denison (eds) *A history of the English language* (pp. 1-42.) Cambridge: Cambridge University Press.
- 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>
- Font, N. (2001). *Rôle de la langue dans l'accès au lexique chez les bilingues: Influence de la proximité orthographique et sémantique interlangue sur la reconnaissance visuelle de mots* (Unpublished Doctoral thesis). Université Paul Valéry, Montpellier.
- Hancin-Bhatt, B., & Nagy, W. (1994). Lexical transfer and second language morphological development. *Applied Psycholinguistics*, 15(3), 289-310. <https://doi.org/10.1017/S0142716400065905>
- Hardie, A. (2012). CQPweb - Combining power, flexibility and usability in a corpus analysis tool. *International Journal of Linguistics*, 17(3), 380-409. <https://doi.org/10.1075/ijcl.17.3.04har>
- Hipfner-Boucher, K., Pasquarella, A., Chen, X., & Deacon, S. H. (2016). Cognate awareness in French immersion students: Contributions to grade 2 reading comprehension. *Scientific Studies of Reading*, 20(5), 389-400. <https://doi.org/10.1080/10888438.2016.1213265>
- Lublinter, S., & Hiebert, E. H. (2011). An analysis of English-Spanish cognates as a source of general academic language. *Bilingual Research Journal*, 34(1), 76-93. <https://doi.org/10.1080/15235882.2011.568589>
- Marín, M. J., & Fernández, P. (2015). The influence of cognates on the acquisition of legal terminology: Help or hindrance? a corpus-based study. *Procedia, Social and Behavioral Sciences*, 198, 320-329. <https://doi.org/10.1016/j.sbspro.2015.07.450>
- Odlin, T. (1989) *Language transfer: cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.
- Otwinowska, A. (2016). *Cognate vocabulary in language acquisition and use: Attitudes, awareness, activation* (Vol. 93). Multilingual Matters.
- Otwinowska, A., & Szewczyk, J. M. (2019). The more similar the better? Factors in learning cognates, false cognates and non-cognate words. *International Journal of Bilingual Education and Bilingualism*, 22(8), 974-991. <https://doi.org/10.1080/13670050.2017.1325834>
- Proctor, C. P., & Mo, E. (2009). The relationship between cognate awareness and English comprehension among Spanish-English bilingual fourth grade students. *TESOL Quarterly*, 43(1), 126-136. <https://doi.org/10.1002/j.1545-7249.2009.tb00232.x>
- 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://doi.org/10.1162/tacl\\_a\\_00024](https://doi.org/10.1162/tacl_a_00024)
- Richards, J. C., & Schmidt, R. W. (2002). *Longman dictionary of language teaching and applied linguistics* (3rd ed.). Longman.
- Ringbom, H. (1987). *The role of the first language in foreign language learning*. Clevedon: Multilingual Matters.
- Ringbom, H. (2001) Lexical transfer in L3 production. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-Linguistic influence in third language acquisition: Psycholinguistic perspectives* (pp. 59-68). Clevedon: Multilingual Matters.
- Schmitt, N. (1997) Vocabulary learning strategies. In N. Schmitt and M. McCarthy (eds) *Vocabulary description, acquisition and pedagogy* (pp.199-227). Cambridge: Cambridge University Press.
- Van Assche, E., Duyck, W., & Brysbaert, M. (2013). Verb processing by bilinguals in sentence contexts: The effect of cognate status and verb tense. *Studies in Second Language Acquisition*, 35(2), 237-259. <https://doi.org/10.1017/S0272263112000873>
- Van Gelderen, E. (2006). *A history of the English language*. Philadelphia: John Benjamins.
- Whitley, M. S. (2002). *Spanish-English contrasts: A course in Spanish linguistics*. Washington, DC: Georgetown University Press.

Winford, D. (2003) *An introduction to contact linguistics*. Oxford: Blackwell Publishing.

Zhou, Z. (2003). *Français-anglais comparaison–Vocabulaire de français universitaire niveau I-IV accéléré surhumaniement* [French-English comparison –Rapid memorization of university French level I–IV vocabulary]. Nanjing: Southeast University Press.

## Appendix 1

*50 cognate pairs randomly selected*

Neighbour cognates (English-French)	Neighbour cognates (English-French)
abundance-abondance	actual-actuel
<b>address<sup>5</sup>-adresse</b>	appeal-appel
<b>board-bord</b>	bomb-bombe
calm-calme	<b>change-changement</b>
chief-chef	<b>class-classe</b>
<b>combatant-combattant</b>	<b>comfortable-confortable</b>
congress-congrès	<b>company-compagnie</b>
comparison-comparaison	<b>complex-complex</b>
<b>consumption-consommation</b>	damage-dommage
<b>department-département</b>	<b>development-développement</b>
<b>domain-domaine</b>	<b>efficacy-efficacité</b>
elementary-élémentaire	<b>enemy-ennemi</b>
<b>example-exemple</b>	formula-formule
genu-genou	<b>group-groupe</b>
inconvenience-inconvénient	<b>independent-indépendant</b>
<b>literature-littérature</b>	maternal-maternel
mathematical-mathématique	<b>method-methode</b>
musician-musicien	Parisian-Parisien
<b>passport-passeport</b>	perfume parfum
<b>personal-personnel</b>	<b>process-processus</b>
pronunciation-prononciation	quotidian quotidien
<b>republic-république</b>	resource-ressource
<b>sympathetic-sympathique</b>	technician-technicien
<b>tissue-tissu</b>	<b>treatment-traitement</b>
vapour-vapeur	vendor-vendeur

<sup>5</sup> Words in bold are those conform to the inclusion criteria included in the result and analysis part.

## Appendix 2

### *Details of query for each neighbour cognates pair*

Neighbour Cognates	Query	Variation	Frequency	Error type
address-adresse	{ad*res*}	address	7	3
		addresses	2	4 (1/3)
		addressed	1	4 (1/3)
board-bord	{bo*rd}	bord	1	1
combatant-combattant	{comba*ant}	combattants	1	2
change-changement	{change*/n}	changement	3	1
class-classe	{class*/n}	classe	2	1
company-compagnie	{compa*n*}	compagnie	9	1
complex-complexe	{complex*}	complexe	1	1
consumption-consommation	{cons*m*tion}	consommation	1	1
		consumation	1	2
comfortable-confortable	{co*fortable}	confortable	4	1
department-département	{depart*ment}	département	8	1
development-développement	{develo*ment}	développement	5	1
		developpment	1	3
domain-domaine	{domain*}	domaine	2	1
		domaines	2	1
example-exemple	{ex*mple}	exemple	7	1
		exemples	1	2
efficacy-efficacité	{effica*/n}	efficacy	2	2
enemy-ennemi	{en*em*/n}	ennemies	2	2
group-groupe	{group*}	groupe	1	1
independent-indépendant	{independ*nt}	indépendant	9	1
literature-littérature	{lit*erature}	littérature	1	1
method-methode	{method*}	methode	2	1
personal-personnel	{person*l}	personnel	13	1
		personnal	3	3
process-processus	{process*/n}	processus	1	1
passport-passeport	{pass*port}	passeport	2	1
republic-république	{republi*}	republique	1	1
sympathetic-sympathique	{sympath*}	sympathic	2	2
treatment-traitement	{tr*t*ment}	traitement	1	1
tissue-tissu	{tissu*}	tissu	1	1