Help or hindrance? The influence of Dutch on the acquisition of English phrasal verbs*

Thomas B. Tienkamp
ACLC, University of Amsterdam

Research has shown that English as a second language (ESL) learners find phrasal verbs (PVs) hard to acquire and often avoid these constructions. An often-neglected aspect in avoidance literature is the uncertainty of knowing the correct verb + particle pairing. The present paper investigated this by looking at Dutch ESL learners' knowledge of verb + particle pairings as well as the influence of the L1 on L2 lexical development. To this end, over 50 Dutch learners of English at two proficiency levels were tested and compared to native speakers on English PVs that either (mis)match in form and meaning as compared to the Dutch counterpart. Results indicate that learners make more mistakes when the L1 and L2 differ in form and meaning. L1 interference is observed as learners overextend Dutch versions into English. This interference need not fossilise as advanced learners perform more like native speakers. Directions for future inquiries are provided as well.

1 Introduction

Inquiries targeting the understanding of the bilingual lexicon typically investigate how L2 learners access the meaning or form of L2 lexical items. A case where syntax and lexicon might come together is the phrasal verb. Phrasal verbs (henceforth: PVs) are defined in this paper as a verb + particle combination in which both elements give up meaning to form one new lexical element (e.g. give up) (Darwin & Gray 1999). Second Language Acquisition (SLA) research has pointed out the difficulty PVs pose for L2 learners. It has been suggested that this difficulty is in part due to them being a peculiarity of the Germanic languages, and in part due to their idiomatic and polysemous meanings (White 2012). A seminal study by Dagut & Laufer (1985) found that Hebrew learners of English avoided phrasal verbs constructions; learners opted for one-word verbs instead. A

* I am indebted to Michelle Nekeman for her help in recruiting participants. I am also thankful to Andrew Mckeown, Penelope Goemans, Parker van Nostrand, and Gautam Ottur for helping me create the stimuli. Lastly, I am grateful for Eva van Lier's and Jan Hulstijn's valuable comments on earlier versions of this paper.

follow up study by Hulstijn & Marchena (1989) showed that Dutch ESL learners do not avoid PVs categorically, but only if they perceive the PV too Dutch-like. In this case, too, Dutch-like was defined as a literal translation of the Dutch form (e.g. *opgeven → give up*). However, Hulstijn & Marchena (1989) did not focus on whether the learner knew the right verb + particle combination. This raises two questions. First, it raises the question as to whether the avoidance can in part be explained due to the learner's uncertainty about the right verb + particle combination. Second, the observation that learners deemed the construction too Dutch-like raises the question of whether typological similarity aids or hinders PV acquisition. To answer these questions, over 50 Dutch ESL learners were tested on PVs that can be directly translated to English (e.g. *opgeven → give up*) and PVs that have different verb + particle combinations (*richten op → aim at*).

The paper is structured as follows. Section 2 introduces Jiang's model of the bilingual lexicon (2000) and relevant studies; Sections 3 and 4 introduce the present study and its methodology; section 5 presents the results; section 6 provides discussion; and section 7 concludes the paper and also provides directions for future inquiries.

2 Theoretical background

2.1 Theoretical models

The Declarative/Procedural model of language (DP model) posits that language is subserved by two memory systems (Ullman 2001, 2004). The declarative memory system, rooted in temporal lobe structures, subserves the lexicon, including the meaning and sounds of words, and idiosyncratic facets of language (e.g. irregular verbs and plural forms). The procedural system, on the other hand, which is rooted in frontal/basal ganglia structures, is important in the learning of habits, computation, and rule-governed aspects of language (e.g. the grammar). The model posits that late learners rely more heavily on the declarative memory system than on the procedural system as the latter may be subject to critical period effects. The former, on the other hand, is not.

Lexical development thus takes place within the declarative memory system. This paper follows de Bot (1992) in assuming that phrasal verbs are lexical in nature since lemmas are activated based on the meaning of the preverbal message. This preverbal message has to then be encoded into words drawn from the lexicon.

Lowie & Verspoor (2004) note that a question relevant to the bilingual mental lexicon is whether and to what extent L2 learners make use of the lexical knowledge from their L1 in the acquisition and use of the L2. Jiang (2000) provides an interactive model of the bilingual lexicon and postulates that there are

three separate stages: (1) the formal stage, (2) the L1 lemma mediation stage, and (3) the L2 integration stage (illustrated in Figure 1, taken from Jiang 2000: 54).

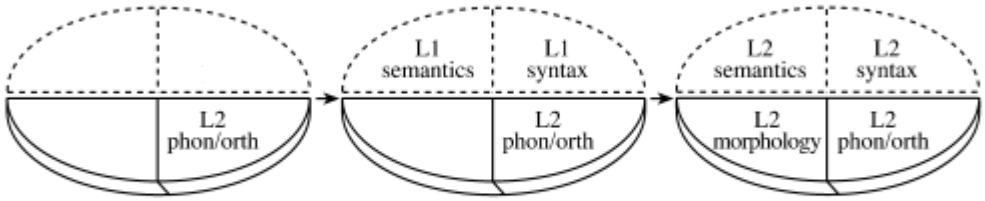

Figure 1: Lexical development from the formal stage (left) to the integration stage (right).

It is important to note the difference between lexemes and lemmas. The lemma contains the semantic and syntactic information (the top level in Figure 1) whereas the lexeme contains the formal and morphological information (bottom level in Figure 1). To illustrate this, the word forms (lexemes) 'give up', 'giving up', and 'gave up' are all part of the overarching lemma GIVE UP.

In the formal stage, a lexeme with formal specifications is established. This lexeme is mapped onto an L1 meaning and there is no direct link from the L2 lexeme to the preverbal message. Then, in the lemma mediation stage, the semantic and syntactic information from the corresponding L1 lemma is copied onto the L2 lemma. The L1 thus mediates a link between the L2 lexical item and the preverbal message. The learner uses their L1 semantic and syntactic knowledge to use the L2 lemma which can lead to transfer mistakes. This stage differs from the formal stage because there *is* a lemma. In the formal stage, there is no information at the lemma level, only at the lexeme level. After sufficient exposure and experience, the L2 information is integrated on the L2 lemma. Now, the L1 information is no longer used and a direct link between the preverbal message and the L2 lemma exists. Once the learner has done so, they have native-like knowledge of this particular lemma.

Jiang (2000) posits that fossilisation of lexical development is indicated by the learner not advancing beyond the second stage. Furthermore, it is postulated that many lexical items may fossilise due to the existence of an L1 lexical and conceptual system. Lowie & Verspoor (2004) also note how this model allows each lexical item to be in a different developmental stage. For example, word A might be fully developed, whereas word B still remains in the lemma mediation stage.

2.2 Relevant studies in PV acquisition

As mentioned in the introduction, most of the research on PV acquisition has focused on avoidance. In this line of research, avoidance was operationalised as the learner's preference for the PV or one-word-verb equivalent in multiple choice

or translation tasks (Dagut & Laufer 1985; Hulstijn & Marchena 1989; Laufer & Eliasson 1993; Liao & Fukuya 2004). Results indicated that learners with a higher proficiency showed less avoidance than learners with a lower proficiency level. There have also been large scale corpus analyses investigating the rate at which learners use PVs as compared to natives. Siyanova & Schmitt (2007) analysed spoken and written corpora of native speakers and ESL learners on their use of PVs. Their results suggest that ESL learners use the one-word-verb equivalent at higher rates than the native speakers do in both spoken and written modalities. This finding supports the previously established avoidance patterns in experimental research as the results mirror each other.

Besides proficiency, typological distance has also been noted as a reason for avoidance. Learners with a Germanic language as their L1 (e.g. Dutch or Swedish) showed considerably less avoidance than learners with a non-Germanic language as their L1 (e.g. Hebrew) (Dagut & Laufer 1985; Hulstijn & Marchena 1989). This already suggests crosslinguistic influence. Another indication of L1 influence comes from the observation by Hulstijn & Marchena (1989) that Dutch learners avoid PVs only if they perceive them to be too Dutch-like (e.g. *opgeven* → *to give up*). These 'too Dutch-like' items are cognates, words whose form and meaning considerably overlap across two languages (Otwinowska & Szewczyk 2019). Psycholinguistic research has found that cognates are retrieved, translated, and recognised faster than non-cognates (Friel & Kenison 2001; Jacobs, Fricke & Krol 2016; Lemhofer & Dijkstra 2004). Furthermore, cognates are more resistant to forgetting, suggesting that they might be easier to learn (De Groot & Keijzer 2000). Indeed, studies have found that cognates are easier to learn than non-cognates (cf. Otwinowska & Szewczyk 2019; for a summary, see De Groot 2011). However, experimental data directly investigating this L1 influence on PV acquisition is currently missing.

L1 influence and/or interference has been noted in other lines of research, too. Lowie & Verspoor (2004) analysed Dutch ESL learners on their knowledge of prepositions. In their experimental material, they used prepositions that were either similar or different in meaning compared to their Dutch counterpart, and used prepositions with varying degrees of frequency. Their results showed that up until high proficiency (first year university students), Dutch interfered with their usage of English prepositions. Learners overextended the Dutch preposition meaning to English resulting in errors like 'I am in the train' instead of 'on the train'. A main effect of frequency was found, too. Higher frequency prepositions were performed better on than lower frequency prepositions. Similar results have been reported by Ijaz (1986), who found that advanced adult ESL learners differed significantly from native speakers with regards to the semantic boundary they ascribed to spatial prepositions. Crucially, the semantic boundary was influenced by the L1 as ESL learners based the L2 semantic boundary on their L1.

3 The present study: Research questions and predictions

On the basis of this background section, it becomes apparent that direct experimental data regarding the L1 influence on the acquisition of PVs is missing. Furthermore, data regarding the learner's knowledge of correct verb + particle combinations is missing as well. Although previous research has suggested that avoidance may also signal this uncertainty, it cannot be taken as fact as Hulstijn & Marchena (1989) did not directly include this in their experimental design. The present study will fill these gaps. The research questions can thus be stated as follows: to what extent does crosslinguistic (dis)similarity influence the acquisition of English phrasal verbs? Do learners reach native-like levels? Do phrasal verbs fossilise in L2 speech? Lastly, what is the role of frequency in PV acquisition? To this end, two groups of Dutch learners of English were tested on PVs that can be directly translated to English, called match items (e.g. *opgeven* → *give up*) and PVs that have different verb + particle combinations, called mismatch items (*richten op* → *aim at*).

Based on the model of Jiang (2000), the results of Lowie & Verspoor (2004), and the cognate literature (De Groot 2011), it is predicted that match items will be easier for learners than mismatch items. Initially, the L1 lemma is copied onto the L2 lemma. As a corollary, it is predicted that mismatch items will pose greater difficulty. This prediction thus goes against Hulstijn & Marchena (1989), who found that match items resulted in more avoidance.

Also based on the model of Jiang (2000), as well as the results of experimental research, it is predicted that advanced learners will score higher than intermediate learners. The advanced learners have received more input and have more experience with the target language. These aspects will help the lemma move from the lemma mediation stage to the lemma integrations stage.

Based on the DP model of language (Ullman 2001, 2004), it is predicted that learners do reach native-like levels and perform more like native speakers than high school students. No decrease in learning ability of the declarative memory system, the system that subserves the lexicon, has been noted and PVs are assumed to be lexical in nature (de Bot 1992). Furthermore, it is predicted that the initial L1 interference will not fossilise due to the absence of a decrease in learning ability. This prediction is thus contra Jiang (2000), who predicts that many lexical items may fossilise.

Lastly, based on Lowie & Verspoor (2004), it is predicted that participants will score higher on higher frequency items than lower frequency items.

4 Methods

4.1 Participants

Three groups of participants were recruited online for the present study: (1) Dutch high school students in their fifth or sixth year of pre-university education (VWO); (2) Dutch university students who have followed an English taught degree for at least three years; and (3) a group of native speakers of English. All but one speaker was currently enrolled at a university or had received a university degree in the past. Following Hulstijn & Marchena (1989) and Lowie & Verspoor (2004), it is assumed that the high school learners resemble intermediate learners, and university students advanced learners. In total, 68 participants were recruited. Group distribution and details can be found in Table 1, where AoA stands for age of acquisition. Participants were recruited via Facebook advertisements, as well as contacts at high schools in Leek and Groningen, The Netherlands.

Table 1: Participant details.

Group	N	Age in Years (SD) Range	AoA in Years (SD) Range	Proficiency
High School	31	M = 17.1 (0.7) R =16–19	M = 9.3 (2.1) R = 6–12	Intermediate
University	23	M = 22.8 (2.2) R = 20–28	M = 9.0 (2.1) R = 6–12	Advanced
Native	14[1]	M = 27.14 (7.6) R = 20–45	N/A	Native

4.2 Materials

4.2.1 LEAP-Q questionnaire

A selection of the Dutch version of the LEAP-Q questionnaire [2] (Marian, Blumenfeld & Kaushanskaya 2007) was imported into Qualtrics (Qualtrics 2019). The LEAP-Q questionnaire provides detailed information regarding the participant's language background, such as age of acquisition (AoA), years of education (YoE), and the exposure to the languages the participant speaks. The included questions can be found in Appendix A.

[1] Six were American English speakers, five were Australian English speakers, two were English English speakers, and one was a Scottish English speaker.

[2] Translated by Lisa Vandeberg, Erasmus University, adapted by Sarah von Grebmer zu Wolfsthurn, Niels Schiller, and Leticia Pablos Robles, Leiden University.

4.2.2 Phrasal verb test

A phrasal verb test consisting of 30 items was created in Qualtrics and aimed to test the participant's knowledge of the correct verb + particle pairing of English phrasal verbs. The influence of the L1 was tested by 15 match items, in which the verb + particle can be literally translated from Dutch to English (e.g. *opgeven* → *to give up*), and 15 mismatch items where the English construction does not mirror the Dutch one (*ophangen* → **hang up*, but *put up*). Both conditions were matched for word frequency. Word frequency was calculated using the Corpus of Contemporary American English (COCA; Davies 2008) using the same method as Liu (2011). This was done by entering the verb lemma plus the particle. The second step was a search of the tokens of transitive PVs used with their particles separated by one or two intervening words. No queries were run for verbs that were separated from their particle by three or more words, as Gardner & Davies (2007) had noted that this results in many false PVs. A complete list of the stimuli plus their frequency per million words can be found in Appendix B. The average frequency per condition is given in Table 2.

Table 2: Average frequency of the stimuli per condition.

Condition	**Frequency per million words (SD)**
Match	22.25 (31.27)
Mismatch	24.17 (37.77)

Participants were given a sentence with a blank and two options regarding verb and particle choices. An example is provided in (1).

(1) Denice is _____ posters all over town for the party she's organising.

A. putting up
B. putting on
C. hanging up
D. hanging on

In the end, 16 match stimuli and 13 mismatch stimuli were included in the analysis. One stimulus turned out not to be a mismatch, and was moved to the match stimuli. A second item (*chase up*) was excluded as native speakers indicated that three out of the four options were acceptable. Stimuli were presented in six fixed blocks of five questions. Blocks were presented in a randomised order and items within blocks were randomised as well. Lastly, all the possible answers were presented in a randomised order as well to prevent any ordering effects.

4.3 Procedures

All participants participated online at a time of their choosing and gave their informed consent. They were first presented with the LEAP-Q questionnaire (approx. 5 minutes) and then with the PV-test (approx. 10–15 minutes). Participants were not rewarded with the correct answers. The answers of the native speakers were assumed to be the correct answers to the questions. In cases where more than 10% of the natives chose for an unintended answer, it was marked as correct as well. This led to the exclusion of the item *chase up* as natives were almost equally divided over three possible answers. There were no ambiguities regarding the correct answer for all other items. The tests of the ESL learners were marked afterwards. Their answers were scored as correct or incorrect. Additionally, when the answer was incorrect, it was indicated whether the mistake was transfer-based or non-transfer based. Transfer mistakes were regarded as literal translations of mismatch items (e.g. *hang up* instead of *put up*). In this example, the transfer results from an overextension of the lexical verb. However, transfer could also result from an overextension of the particle (e.g. *come over as* instead of *come off as*).

4.4 Analyses

All statistics were carried out using R and the lme4 and lmerTest packages (Bates, Maechler, Bolker & Walker 2015; Kuznetsova, Brockhoff & Christensen 2017; R-core team 2018). In order to test whether there were group differences for the reported scores for proficiency and exposure to English, independent samples *t*-test were run with group as the independent variable and reported proficiency and exposure as the dependent variable.

In order to test where advanced learners are as compared to intermediate learners and natives on the PV-test, a linear mixed effects model was fitted with percentage correct as a function of the ternary predictor group (High school, University, and Native), and the binary predictor condition (Match and Mismatch).[3] A second linear mixed effect model was fitted with percentage correct as a function of the binary predictors group (High school and University) and condition (Match and Mismatch) to investigate group differences. Categorical predictors were coded with sum-to-zero orthogonal contrasts. A linear mixed effects model was opted for instead of a generalised linear mixed effects model due to the difference in condition sizes (13 items in mismatch, 16 items in match).

[3] Other models were run as well with predictors from the LEAP-Q questionnaire. These predictors included: age of acquisition, years of education, reported proficiency, reported exposure. None of these predictors improved the models, as predictors were strongly correlated. They were therefore left out.

The statistic that will answer the research question is the interaction between group and condition.

To still analyse the effect of frequency, as this was not possible using aggregated scores in the linear model, a generalised mixed effects model was fitted with correct as a function of the binary predictors group (High school and University) and condition (Match and Mismatch), and the continuous predictor of frequency per million words

Next, an error analysis was carried out. An independent samples *t*-test was run to check for a group difference in amount of transfer mistakes. Lastly, two paired *t*-tests were carried out, one for each group, to test whether participants made more transfer mistakes than non-transfer mistakes.

5 Results

Descriptive results of both the LEAP-Q questionnaire and the PV-test can be found in Table 3. These results (in percentages) are visualised in a boxplot in Figure 2.

Table 3: Descriptive statistics Experiment 1.

Group	Exposure to English (in %)	Proficiency (From 1-10)	Total score (Max 29)	Match score (Max 16)	Mismatch score (Max 13)
High School	24.84 (16.2)	7.16 (0.94)	20.5 (5.2)	12.9 (2.8)	7.6 (2.8)
University	39.9 (14.9)	8.97 (1.67)	27.6 (1.1)	15.7 (0.57)	12.0 (1.0)
Native	N/A	N/A	28.7 (0.61)	15.86 (0.53)	12.86 (0.36)

The independent samples *t*-tests investigating group differences in reported exposure and proficiency were both significant. The university students in the experiment reported significantly higher English proficiency than the high school students in the experiment (estimated group difference = 1.8 points; 95% CI = 1.26…2.36 points; $p < 0.001$). They also reported significantly higher exposure to English than the high school students (estimated group difference = 14.86%; 95% CI = 8.89…20.83%; $p < 0.001$).

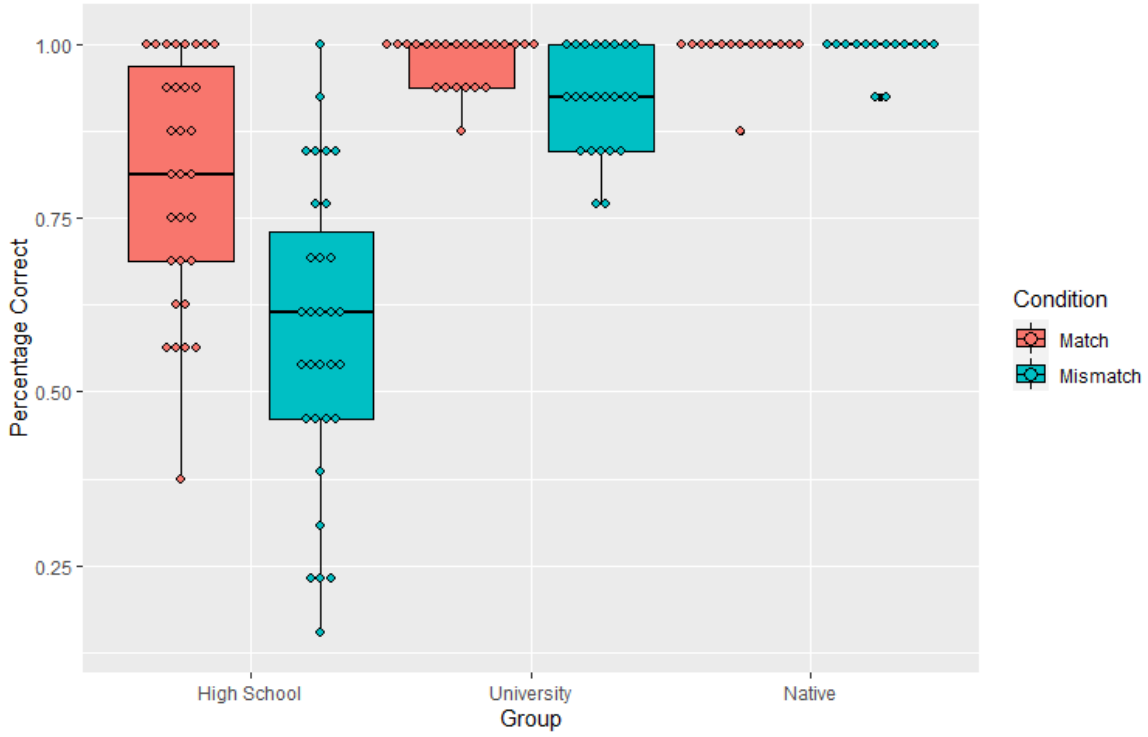

Figure 2: Percentage correct on the PV-test across conditions and groups. Dots represent the individual scores of participants.

The linear mixed effects model investigating whether advanced learners are more like native speakers than intermediate learners was significant. Advanced learners scored significantly, on average, 10.66% higher than the average of native speakers and intermediate learners combined (95% CI = 4.26%…17.06%; p = 0.002). It is therefore concluded that advanced learners are more like native speakers than intermediate learners. As a sanity check, it was checked whether native speakers performed better than intermediate learners. Native speakers performed significantly better than intermediate learners (estimated group difference: 29.5%; 95% CI: 21.7%…37.3%; $p < 0.001$).

The linear mixed effects model investigating group differences on the PV-test was also significant. There were significant effects for group, condition, and a significant interaction between group and condition. The university group scored significantly higher than the high school group (estimated group difference = 25.42%; 95% CI = 17.91%…32.93%; $p < 0.001$). Participants averaged across groups scored significantly higher on match trials than mismatch trials (estimated condition difference = 14.09%; 95% CI = 10.45%…17.74%; $p < 0.001$). The interaction indicated that high school students scored an additional 16.48% lower on mismatch trials than on match trials when compared to university students (95% CI = 9.19%…23.77%; $p < 0.001$). It is therefore concluded that the group difference is larger in the mismatch condition than in the match condition.

As for the effect of frequency, the generalised mixed effects model showed a significant main effect. Participants across groups were 1.033 times more likely to pick the correct verb + particle combination on PVs with a higher frequency than PVs with a lower frequency (95% CI = 1.003…1.064 times more likely; $p < 0.05$).[4]

The descriptive statistics for the error analysis can be found in Table 4 and is visualised in a boxplot (in percentages) in Figure 3. There was a significant group difference. High school students made 2.4 more transfer mistakes on average as compared to university students (95% CI = 1.48…3.32 points; $p < 0.001$).

Table 4: Error distribution of the learner groups.

Group	Transfer mistakes	Non-transfer mistakes
High School	3.55 (1.70)	2.07 (1.81)
University	1.13 (0.64)	0.4 (0.51)

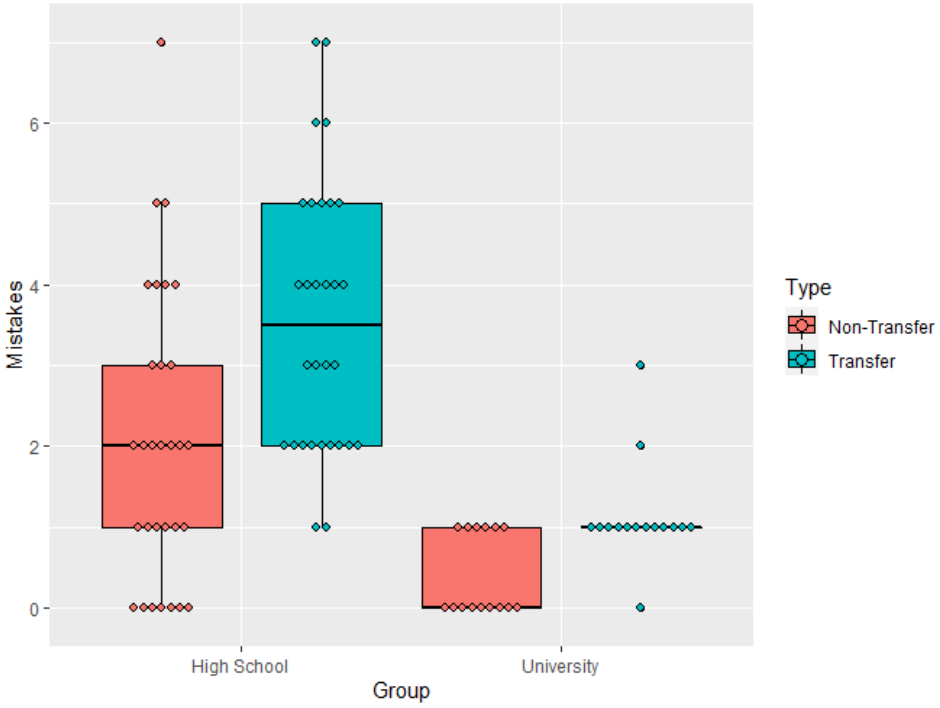

Figure 3: Error distribution of the learner groups.
Dots represent the individual scores of participants.

[4] Generalised mixed effect models yield a log-odds outcome. Therefore, the outcome was exponentionalised and interpreted as X- times more/less likely than.

Paired *t*-tests indicated that both groups were more likely to make transfer mistakes than non-transfer mistakes. High school students made an estimated 1.47 more transfer mistakes than non-transfer mistakes (95% CI = 0.62…2.31 mistakes; $p < 0.001$). University students made an estimated 0.73 more transfer mistakes than non-transfer mistakes (95% CI = 0.24…1.22 mistakes; $p < 0.001$).

6 Discussion

The aim of the present study was to answer two questions. The main question was whether typological similarity aids or hinders PV acquisition. This question was addressed from a developmental perspective by looking at intermediate learners (high school students) and advanced learners (university students). Based on the model of Jiang (2000) and the DP model of language (Ullman 2001, 2004), it was predicted that (1) university students should perform better than high school students in both conditions, (2) that mismatch items should pose the largest difficulty for both groups, (3) that there should be more transfer than non-transfer mistakes, (4) that lexical overextension need not fossilise provided that the learner receives enough input, and (5) that learners would perform better on higher frequency items than lower frequency items.

Differences were found in overall ability to correctly pick the right verb + particle pairing, with university students performing better than the high school students, and participants from both groups performing better on match trials than mismatch trials. There was also considerably less variation in the scores of university students as compared to high school students as evidenced by the small standard deviations. Additionally, participants from both groups were more likely to overextend a Dutch form into English than making a different kind of mistake in mismatch trials. These results confirm predictions (1–3). It can be concluded that L1 Dutch hinders the acquisition of L2 PVs where there is a form meaning mismatch. The error analysis and the discrepancy between match and mismatch trials suggest that learners assume, albeit unconsciously, that the semantics of the L1 and L2 forms align. The results thus support the notion of lemma mediation in Jiang's (2000) model. Input and experience is needed to learn and integrate the L2 semantic and syntactic information for the L2 lemma. This is further reflected by the significant effect of frequency. Learners have more experience with and come across higher frequency items more often, thus having more opportunity to integrate the L2 specific information. This main finding, albeit a small one, supports prediction (5).

Lemma mediation need not be a hindrance in all cases, though. Dutch seems to aid PV acquisition in cases where form and meaning match. In these instances, integration is not necessary as target-like production does not depend on L2 lemma integration. This is an interesting finding, as Hulstijn & Marchena (1989)

argued that Dutch ESL learners refrain from using PV constructions that have identical Dutch counterparts because they are too 'Dutch-like'. So, while L1-L2 similarity might hinder the actual use of the construction, the results of the present study suggest that similarity is helpful in the acquisition stage. The study thus supports the prediction made by De Groot & Keijzer (2000) that cognates might be easier to learn.

Results are also in line with previous studies that found that groups with a higher proficiency performed more native-like (Hulstijn & Marchena 1989; Liao & Fukuta 2004; Lowie & Verspoor 2004). Furthermore, Lowie & Verpoor (2004) reported a similar discrepancy between form meaning (mis)matches in L2 prepositions for Dutch learners of English. They reported higher accuracy when there was L1-L2 similarity. This effect was not found for the group with the highest proficiency, who performed well in both conditions. The present study mirrors this finding with the significant interaction between group and condition.

The results also provide support for prediction (4). The significant interaction between group and condition suggests that advanced learners managed to bridge the initial gap between conditions. Advanced learners performed almost equally well on both conditions. Jiang (2000) predicts that many lexical items fossilise, meaning that the learner remains in stage two, the lemma mediation stage. This prediction was not borne out by the data in the present study. The DP model (Ullman 2001, 2004), on the other hand, posits that the lexicon is subserved in the declarative memory system. This memory system is not constrained by maturational constraints like the procedural system that subserves grammar and computation. Thus, the DP model predicts that there may be L1-L2 differences in the level of attainment of grammar, but not necessarily in the lexicon. The results of the present study support the DP model in this aspect as the advanced learners performed significantly more like native speakers than intermediate learners. However, it should be noted that the few mistakes advanced learners made, were mostly transfer-based. It should also be noted that this significant difference was based on a mean difference of only 0.7 as advanced students made, on average, 1.14 transfer mistakes versus 0.4 non-transfer mistakes (see Table 4 and Figure 3). So, although there is no considerable fossilisation in terms of number of mistakes, it is hypothesised that the mistakes that do remain show that the learner is still in stage 2 for those specific items. This also reflects the flexibility of Jiang's (2000) model, as different lexical items may be in different developmental stages.

It has been suggested in the literature that compositional PVs, where the meaning of the PV equals the combined meaning of its components, are built online (González 2010). In this instance, the PV would not be stored as a whole in the lexicon but put together by means of computation. As computation is subserved by the procedural memory, it could be the case that these items *are*

prone to fossilisation. However, it was beyond the scope of the paper to investigate this more closely. Future inquiries could look into this.

The second question was whether PV avoidance can in part be explained by the uncertainty of knowing the correct verb + particle pairing. As mentioned in the introduction, most PV studies focused on preference, not performance (see Dagut & Laufer 1985; Laufer & Eliasson 1993; Hulstijn & Marchena 1989). They posit that the learners know these constructions, either by a short test of whether the learner understands the PV, or by presupposing their knowledge. However, Jiang (2000) notes that knowing a word(string) need not imply that the learner is also able to use the construction correctly. Very often, learners' receptive vocabulary is larger than their productive vocabulary (Schmitt 2010). The results of the present study indicate that the intermediate learners have difficulty with choosing the right verb + particle pairing, failing to pick the correct pairing in 30% of the cases. This is also the group where significant differences in preference have been reported when compared to native speakers (Hulstijn & Marchena 1989). The advanced learners in Hulstijn & Marchena (1989) did not differ significantly from native speakers in their preference. In the present study, the advanced learners showcased extensive knowledge of the correct verb + particle pairings, scoring over 90% in both conditions. It could thus very well be the case that part of the avoidance that is reported is due to the uncertainty of not knowing the right construction. Although this hypothesis is validated by cross experiment comparison, within participant experimental data is needed in order to accept or reject this hypothesis. Experimental designs of Hulstijn & Marchena (1989) and the present study may be combined in order to do so in future inquiries.

Although the present study has shed light on the influence of the L1 in L2 lexical development, some limitations must be noted, as well. Two kinds of limitations will be discussed: (i) design limitations and (ii) data limitations.

As for the design limitations, the present study consisted of one multiple-choice task in a in a cross-sectional, non-longitudinal design. Previous studies comprised the administration of multiple tasks, thus reducing the chance of task-bias. The results of the present study might have been influenced by the task choice. Furthermore, evidence, or the lack thereof, of fossilisation ideally comes from longitudinal research. However, due to time-constraints, a cross-sectional design was opted for instead.

Secondly, even though the present study improved on previous work by including a proficiency measure, it can still be seen as a limitation. Self-reported proficiency is limited in the sense that it remains doubtful whether non-experts can correctly assess their proficiency. It is therefore recommended that future studies employ an objective measure of proficiency and use multiple tasks using a longitudinal design that probe both language knowledge and use.

As for the data limitations, there was a lack of experimental control as the data was collected online. It could have been the case that participants (1) looked up the right answers or (2) did not try and filled in random answers. Although the first cannot be entirely ruled out, the results do indicate that participants filled in the test in earnest. As for latter, participation was voluntary and therefore seems unlikely. Additionally, the time it took the participants to fill in the online test did not suggest that participants rushed through it. The time further suggests that participants did not look up the right answers, as this would have increased their time significantly.

The second data limitation is the frequency data. Some of the used verbs were polysemous, and the corpus that was used is not tagged for semantic content. Moon (1998) suggests that around five percent of phrasal lexemes are polysemous. Thus, the attained frequency numbers are for all the different meanings that each lexeme has. This means that the frequencies reported in Appendix A do not correspond one-to-one with the meaning of the PV, but rather give an indication of the verb's frequency in general. However, this limitation is only secondary, as frequency was not of primary interest in the present study.

7 Conclusion

This paper has made a contribution to the avoidance literature on PVs by examining the learner's knowledge of the correct verb + particle pairing. Additionally, this paper examined the role of the L1 in L2 PV acquisition. It has been argued that the L1 may be of help when L1 and L2 align in form-meaning mappings. When they do not, the L1 may be a hindrance. The results reflect this as participants scored higher on match items than mismatch items. However, the results also indicate that this hindrance need not be long term. Initial transfer mistakes do not fossilise in L2 speech. The results can thus be explained by the model of Jiang (2000), who proposed lexical development in terms of L1 lemma mediation and L2 information integration. However, contra Jiang (2000), the present paper showed that lexical items do not fossilise as often as initially proposed. The expected non-fossilisation was motivated by the division of lexicon and grammar in the DP model instead (Ullman 2001, 2004).

Two fruitful directions for future inquiries have been noted in the discussion. First, inquiries may tackle the distinction between compositional and non-compositional PVs in terms of lexical fossilisation. Second, the methodologies of traditional avoidance research and the present study may be combined to directly assess the role of not knowing the verb + particle pairing in avoidance through the administration of multiple tasks that probe language *knowledge* as well as language *use*. This would ideally be investigated in a longitudinal design. Future inquiries will further elucidate why ESL learners find

them so challenging. Understanding the nature of the experienced difficulties will guide the field in helping ESL learners acquire these constructions better and make them more confident in using them.

8 References

Bates, D., M. Maechler, B. Bolker & S. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67: 1–48.

Dagut, M. & B. Laufer. 1985. Avoidance of phrasal verbs – A case for contrastive analysis. *Studies in Second Language Acquisition 7*: 73–79.

Darwin, C.M. & L.S. Gray. 1999. Going after the phrasal verb: An alternative approach to classification. *TESOL Quarterly* 33: 65–83.

Davies, M. 2008. *The Corpus of Contemporary American English (COCA): One billion million words, 1990-2019*. https://www.english-corpora.org/coca/.

de Bot, K. 1992. A bilingual production model: Levelt's 'speaking' model approach. *Applied Linguistics* 13(1): 1–24.

de Groot, A.M.B. 2011. *Language and Cognition in Bilinguals and Multilinguals: An Introduction*. New York: Psychology Press.

de Groot, A.M.B. & R. Keijzer. 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.

Friel, B.M. & S.M. Kennison. 2001. Identifying German-English cognates, false cognates, and non-cognates: Methodological issues and descriptive norms. *Bilingualism: Language and Cognition* 4(3): 249–274.

González, R.A. 2010. Making sense of phrasal verbs: A cognitive linguistic account of L2 learning. *AILA Review* 23(1): 50–71.

Hulstijn, J. & E. Marchena. 1989. Avoidance: Grammatical or semantic causes? *Studies in Second Language Acquisition* 11(3): 241–255.

Ijaz, I.H. 1986. Linguistic and cognitive determinants of lexical acquisition in a second language. *Language Learning* 36(4): 401–451.

Jacobs, A., M. Fricke & J.F. Kroll. 2016. Cross-language activation begins during speech planning and extends into second language speech. *Language Learning* 66: 324–353.

Jiang, N. 2000. Lexical representation and development in a second language. *Applied Linguistics* 21(1): 47–77.

Kuznetsova, A., P.B. Brockhoff & R.H. Christensen. 2017. lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* 82(13): 1–26.

Laufer, B. & S. Eliasson. 1993. What causes avoidance in L2 learning: L1-L2 difference, L1-L2 similarity, or L2 complexity? *Studies in Second Language Acquisition* 15(1): 35–48.

Lemhofer, K. & T. Dijkstra. 2004. Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition* 32: 533–550.

Liao, Y. & Y. Fukuya. 2004. Avoidance of phrasal verbs: The case of Chinese learners of English. *Language Learning* 54(2): 193–226.

Liu, D. 2011. The most frequently used English phrasal verbs in American and British English: a multicorpus examination. *Tesol Quarterly* 45(4): 661–688.

Lowie, W. & M. Verspoor. 2004. Input versus transfer? The role of frequency and similarity in the acquisition of L2 prepositions. In M. Achard & S. Niemeier (eds.), *Cognitive Linguistics, Second Language Acquisition, and Foreign Language Teaching*. Berlin: Mouton de Gruyter, 77–93.

Marian, V., H.K. Blumenfeld & M. Kaushanskaya. 2007. The language experience and proficiency questionnaire (leap-q): assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research* 50(4): 940–67.

Moon, R. 1998. Frequencies and forms of phrasal lexemes in English. In A.P. Cowie (ed.), *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press, 79–100.

Otwinowska, A. & J.M. Szewczyk. 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): 1–18.

Qualtrics. 2019. Version 20.05. Utah, USA. https://www.qualtrics.com.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org.

Schmitt, N. 2010. *Researching Vocabulary: A Vocabulary Research Manual*. London: Palgrave Macmillan.

Siyanova, A. & N. Schmitt. 2007. Native and non-native use of multi-word vs. one-word verbs. *International Review of Applied Linguistics* 45: 119–139.

Ullman, M. 2001. The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism, Language and Cognition* 4: 105–122.

Ullman, M. 2004. Contributions of memory circuits to language: The declarative/procedural model. *Cognition* 92(1): 231–270.

White, B.J. 2012. A conceptual approach to the instruction of phrasal verbs. *Modern Language Journal* 96(3): 419–438.

Thomas Tienkamp
Amsterdam Center for Language and Communication
Spuistraat 134
1012 VB Amsterdam
thomas.tienkamp@student.uva.nl

Appendix A: LEAP-Q questions included in the study:

1. What is your gender?
2. What is your age?
3. How many years of formal education do you have since primary school?
4. What is your current education level?
5. Please list all the languages you know in terms of dominance
6. Please list all the languages you know in terms of order of acquisition
7. Please list what percentage of the time you are currently and on average to all your languages.
8. When you could read a text in your specified languages, in what percentage of cases would you choose each language?
9. When you could speak to a person who speaks all languages equally well as you, in what percentage of cases would you choose each language?

The following questions were about English specifically

1. At what age did you start learning English?
2. Please specify your knowledge of English on a 1-10 scale on speaking, reading, and understanding English
3. On a scale from 1-10, please select how much the following factors contributed to you learning English (interaction with friends, interaction with family, reading, (self)instruction, watching tv, music)
4. On a scale from 1-10 please specify to what extent you are currently exposed to English in the following contexts (interaction with friends, interaction with family, reading, (self)instruction, watching tv, music)

Appendix B: Stimuli used in the experiment and their frequency per million words

Match trials

English PV	Dutch translation	Frequency per million words
bite off	afbijten	1.393
count on	rekenen op	13.922
call back	terugbellen	10.881
blow off	afblazen	5.163
carry out	uitvoeren	26.225
give away	weggeven	13.92
shut off	uitdoen	4.492
tear out	uitscheuren	1.768
stick out	uitsteken	8.002
take back	terugnemen	23.622
watch out	uitkijken	8.029
focus on	focussen op	119.999
stand for	staan voor	20.671
come off	afkomen	14.289
try on	aanpassen	9.166
get back	terugkomen	74.4

Mismatch trials

English PV	Dutch translation	Frequency per million words
talk over	doorheen praten	2.355
come off as	overkomen als	14.289
doze off	indutten	0.925
keep to	houden voor	12.785
think of	denken aan	139.136
wander off	weglopen	1.427
be dependent on	afhankelijk zijn van	45.683
aim at	richten op	18.981
jump at	springen op	2.561
put up	ophangen	26.105
be angry with	boos zijn op	1.879
vote for	stemmen op	42.54
strike out	doorstrepen	5.485