Bilingualism: Language and Cognition

cambridge.org/bil

Paying attention to verb-noun collocations among returnees and heritage speakers: How vulnerable are L2 English collocations to attrition?

Hadil Alraddadi[1,2], Fraibet Aveledo[2], Roland Hangelbroek[3] and Jeanine Treffers-Daller[2]

[1]Taibah University, Department of Languages and Translation, Medina, Saudi Arabia; [2]University of Reading, Department of English Language and Applied Linguistics, Reading, UK and [3]Scientific Intelligence, Novo Nordisk A/S, Måløv Denmark

Research Article

Cite this article: Alraddadi, H., Aveledo, F., Hangelbroek, R. and Treffers-Daller, J. (2024). *Paying attention* to verb-noun collocations among returnees and heritage speakers: How vulnerable are L2 English collocations to attrition? *Bilingualism: Language and Cognition*, 1–14
https://doi.org/10.1017/S1366728924000610

Received: 18 January 2024
Revised: 02 July 2024
Accepted: 06 July 2024

Keywords:
language attrition; returnees; heritage speakers; collocations; processing

Corresponding author:
Hadil Alraddadi;
Email: haraddadi@taibahu.edu.sa

I would like to express my sincere gratitude to Taibah University for providing me with the invaluable opportunity to pursue my studies. We would also like to thank the participants for their time and valuable input in this study. We should note that Novo Nordisk was not involved in any aspect of this study.

Abstract

It is well established that verb-noun collocations are difficult for L2 learners, but little is known about the extent to which such collocations are vulnerable to attrition under conditions of reduced input. The study is novel in that we focus on L2 attrition rather than L1 attrition, and because we focus on Saudi Arabian returnees, who have so far hardly been studied. These are compared to child, adolescent and adult heritage speakers in the US. Receptive knowledge of English collocations was measured with a novel online acceptability judgement task and an online gap-filling task. We found that child returnees experienced more difficulties than the adolescent returnees, because the child returnees had not acquired collocations to the same extent as the adolescent returnees, and they experienced more crosslinguistic influence from Arabic. The current study also provides some counter evidence against the claim that every bilingual is an attriter.

Introduction

Anyone who has lost regular contact with a language that has already been acquired is likely to experience changes in proficiency and use of that language, a phenomenon that is often referred to as 'language attrition' (Schmid & Köpke, 2017). Heritage speakers (HSs) are a particularly interesting group of bilinguals for the study of attrition, because they grow up with a first language (their heritage language, HL) while living in an environment where another language is spoken by the majority. This majority language is generally the HSs' second language (L2). While HSs are often dominant in their HL in the first years of their life, in the course of development, they become more dominant in the majority language (Kupisch et al., 2021). While most studies in the field (see Brehmer & Treffers-Daller, 2020, for examples) focus on the attrition of the HL, less is known about L2 attrition among HSs who have returned to their country of origin and have lost daily contact with L2 after return (Flores, 2019). Studying this group, generally called returnees, offers new perspectives on the linguistic features that are vulnerable in attrition. Returnees are a bilingual group who are particularly vulnerable to dominance shift processes (Flores et al., 2022), and they experience what Flores (2019) has called REPEATED INPUT ALTERNATION. That is, bilinguals who have stayed in an L2 setting from early childhood or birth are initially exposed to their community language (L1) and then gradually become immersed in the dominant language (L2). Thus, L2 becomes the bilinguals' dominant language. However, upon returning to their parents' homeland, their L1 becomes dominant once again, while their exposure and fluency in the L2 decreases. Therefore, these returnees experience an L1 and L2 reversal at a certain age during childhood or adolescence.

Prior studies have shown that the degree of attrition that affects the returnees' L2, upon their return, is correlated with the age at which they changed their dominant linguistic environment, which is referred to as the age of return (AoR). Whether return happens during or after childhood is important for returnees' language profiles (Flores, 2010, 2019). Several studies have reported a rapid decline in linguistic competence among child returnees, namely those who no longer have regular L2 exposure during childhood as opposed to returnees who return as adolescents (Flores, 2010). According to Bylund (2019), the process of attrition is likely to be severe during the pre-pubertal period since a child's linguistic knowledge is much more vulnerable to attrition than adults'/adolescents' linguistic knowledge. By contrast, other evidence suggested that AoR had no influence on L2 attrition among returnees and may depend on other variables, (e.g., language exposure, language use and attitudes) (Tomiyama, 2009). This view is supported by Kubota et al. (2020a), who found that a different outcome can be seen among Japanese child returnees who continue to have regular L2 contact, for example because they receive formal L2 instruction. The different retention rates of L2 grammar between child and adolescent returnees have led Flores

(2010) to suggest that developing a native language involves two distinct processes: the acquisition of linguistic knowledge followed by a stabilization period. Based on neurocognitive approaches to language attrition, Kubota et al. (2020b) propose that most domains of a native language become relatively resistant to disuse effects once consolidated. However, it is not sufficiently investigated which specific language domains and linguistic structures are subject to stabilization, and how long this period may last.

A substantial number of studies on attrition focus on phonology (de Leeuw et al., 2018), morphology (Matos & Flores, 2022), syntax (Flores, 2010; Sá-Leite et al., 2023). However, much less attention has been paid to attrition in the lexicon (Schmid & Jarvis, 2014). Previous studies on lexical attrition mostly focused on single-word units, while knowledge and use of multiword units (MWUs) have hardly been investigated (Kopotev et al., 2020). The latter suggest that HSs use transfer-based non-standard word combinations, and that analysing such combinations can throw new light on the role of input in HSs' language development. Regardless of the specific features under study, the extent and rate of attrition can vary depending on extralinguistic factors, such as age at immigration, length of residence (LoR), and exposure to both languages (Mehotcheva & Mytara, 2019).

Evidence for attrition in MWUs can be found by comparing returnees against HSs. Returnees and HSs are comparable because the returnees experienced the same language development as HSs until the time of return to the country of the home language (Treffers-Daller et al., 2016). Thus, HSs can become a source of information about the language profiles returnees would have had at the moment of return. Specifically, it is the change of environment during a crucial period of development (pre- or post-puberty) which makes this returnee group a unique testing ground for examining both theoretical and empirical aspects concerning language attrition and retention processes. Research on this particular group remains scarce, however, partly due to the difficulty of recruiting returnee participants (Matos & Flores, 2022).

The current study focuses on a specific type of MWUs, namely collocations, which are known to be very difficult to acquire for L2 learners, and potentially vulnerable under conditions of reduced input (Pulido, 2022). Collocations refer to expressions such as *strong tea* or *take a picture*, where the components have a syntagmatic relationship (e.g., modifier + head or verb + object) (Wood, 2019). To what extent HSs are successful in learning age appropriate MWUs in their L2 remains largely unknown, and we know even less about the vulnerability of these L2 MWUs to attrition after return. However, if returnees lose daily access to the majority language after return, this might lead to attrition in productive and/or receptive skills in representations or processing of L2 MWUs. Previous literature on word knowledge suggested that mastering MWUs productively takes time, because productive skills tend to emerge later in the learning process (González-Fernández & Schmitt, 2020). In the context of the present study on MWUs, it may pose an additional layer of complexity. The ability to learn patterns, particularly collocational patterns, involving word combinations, may diminish over time (Arnon et al., 2017). Thus, returnees may encounter difficulties maintaining and applying more complex linguistic structures because of reduced L2 exposure.

Whether an L2 MWU is really attrited can be investigated by comparing children who had returned before puberty to those who returned after puberty: Pre-puberty returnees might not have acquired complex lexis in their L1 or L2 prior to return, whereas post-puberty returnees are likely to have consolidated their knowledge of these structures. Thus, among post-puberty returnees, any lack of MWUs knowledge and/or difficulties experienced when processing these structures, is likely the result of attrition, but pre-puberty returnees may not yet have acquired them in the first place.

The key reasons why MWUs are difficult is that, first of all they are partly arbitrary in that it is unclear why it is *reach a decision* and not *meet a decision*, but *meet a deadline*, and not *reach a deadline* (Szudarski, 2012, p.5). Second, verb-noun (VN) collocations are often not congruent between two languages (English *pay attention* translates as *faire attention* "make attention" in French). Unsurprisingly therefore, L2 learners frequently produce novel collocations, many of which are influenced by L1 influence (Laufer and Waldman, 2011). Indeed, as is well-known, the existence of partial overlap between two languages can lead to crosslinguistic influence (CLI), or crosslinguistic overcorrection (CLO) (Kupisch, 2014). CLI refers to situations where the dominant language influences the HL directly, resulting in the speaker using the SAME structures in both languages. However, CLO refers to situations in which this influence has an indirect effect, resulting in a preference for a particular form in the HL that DIFFERS from that of the dominant language (Anderssen et al., 2018). According to Kupisch (2014), bilinguals tend to overstress the *differences* rather than the *similarities* between their two languages. This has been explained as 'over-inhibition' of structures in the dominant language, which also affects similar structures in the HL (Anderssen et al., 2018). Within the syntactic domain, several studies found evidence for the existence of CLI from L1 among returnees (Flores, 2010). Anderssen et al. (2018), by contrast, found evidence for CLO in that HSs tended to overuse structures that differed maximally from English structures. Despite the insights provided by these studies, attrition in the lexical domain among returnees remains understudied.

L2 Collocational Processing

When it comes to L2 processing of collocations, the volume of research is small and empirical studies on the representation and processing of MWUs have only recently emerged in the field (Pulido, 2022). More recently, research has focused on L1–L2 collocation congruency, that is, the possibility of a direct translation match between the two languages. Previous findings revealed that L2 learners process congruent L2 collocations more quickly and more accurately than incongruent collocations (Yamashita & Jiang, 2010). Furthermore, as shown in Laufer and Waldman (2011), using incongruent collocations is challenging even for advanced learners.

There are only a few studies which examined collocational attrition among Arabic-speaking learners of English (e.g., Alharthi, 2015; Zaabalawi, 2019). Alharthi (2015) shows that attrition is more pronounced in productive tasks as opposed to receptive tasks tapping into formulaic language. In addition, previous studies on Arabic learners used untimed, offline collocation tasks, and did not examine learners' collocational processing in real-time production, which is why further research into this area is urgently needed.

The Present Study

Our approach to the study of attrition is novel for the following reasons. Firstly, in contrast to previous bilingualism research that focused primarily on monolingual norms for comparisons, we examine a bilingual reference group consisting of adult HSs of Arabic that are from the same cultural background as the returnees. We avoid the use of a monolingual baseline group, because the appropriateness of a

https://doi.org/10.1017/S1366728924000610 Published online by Cambridge University Press

monolingual comparison group has been queried by many researchers investigating the cognitive and linguistic characteristics of bilinguals (De Houwer, 2023). Thus, we can gain new insights into the acquisition of HLs spoken in immigration contexts and L1 and L2 attrition among returnees by comparing returnees with HSs remaining in the host country, which means comparing two bilingual groups against each other rather than comparing bilinguals against monolinguals (see Flores, 2010; Treffers-Daller et al., 2016). As highlighted by Rothman and Treffers-Daller (2014), HSs are native speakers (NSs) of their languages, too. Moreover, because bilinguals acquire two languages, the possibility of mutual language influence is constantly present, which makes them better controls than monolinguals. Secondly, there are remarkably few studies on L2 attrition among returnees (Flores, 2010, 2019; Kubota et al., 2020a, 2020b; Matos & Flores, 2022; Tomiyama, 2009) and only one published study (Treffers-Daller et al., 2016) examines collocational use among returnees and HSs. There have also been calls for the use of more psycholinguistic techniques in future studies of L2 collocational processing for assessing recall (Sonbul & El-Dakhs, 2020). To the best of our knowledge, no research explores the processing of L2 collocations among returnees. This study employs timed psycholinguistic tasks that are assumed to reflect automatic language processing, in contrast with early studies that used offline tests only.

The specific aim of the current study is to investigate to what extent there is evidence for attrition in the processing of English VN collocations among L1 Arabic-speaking returnees and, if so, whether L1 influence is responsible for any difficulties they experience. VN collocations were chosen because these are well-known to be complex for L2 learners (Boers et al., 2014; Laufer & Waldman, 2011). Receptive knowledge of English collocations was measured with a novel online acceptability judgement task (AJT); an online gap-filling task (GFT) was used to measure productive knowledge of these constructions. We aim to establish to what extent a) returnees underperform by comparison of Saudi Arabian HSs living in the United States (US); and b) if so, whether this is because these had not yet been acquired prior to return, or to attrition. Thus, this study aims to address the following research questions:

Firstly, to be able to explain the presence or absence of attrition, we focus on differences in language dominance.

RQ1: Is there any difference in the relative degree of language dominance among the HSs and the returnees?

All HSs were expected to be English-dominant and the child returnees who returned to Saudi Arabia (SA) before the age of 11, to be Arabic-dominant, based on previous observations on heritage language and L2 development (Montrul, 2016)[1]. The situation might be less straightforward for the adolescent returnees. In light of Flores's (2010) study, it was predicted that the adolescent returnees would be situated between the HSs and the child returnees due to their extended stays in the US and SA, and due to their decreased English exposure and increased Arabic exposure, causing a higher degree of balance. Therefore, the adolescent returnees were predicted to be the most balanced group.

RQ2: To what extent is there evidence for attrition in receptive and productive knowledge of English collocations among adult Arabic-English bilingual returnees living in SA by comparison with adult Saudi HSs living in the US?

The adult HSs (AHS) in the US are needed to provide a baseline for analysing the returnees' collocational knowledge, as the AHS represent the knowledge HSs develop under conditions of continued English input in the US. If the returnees obtain lower scores on collocation tests (and are slower in replying) than the AHS, as might be expected, one possibility is that the returnees knew these collocations at the time of return but lost access to these after return. This would mean the attrition scenario is the most likely one. Alternatively, it is possible they left before they had acquired these, which means the attrition scenario does not apply. To be able to answer the question, we collected data from two groups of returnees: child returnees, who left the US between the ages of five and eleven and adolescent returnees, who left after the age of twelve. These were compared to Arabic L1 child and adolescent HSs who were studied at the ages at which the returnees left the US.

We begin by investigating whether the two returnee groups underperform by comparison with the AHS. Subsequently, we compare the subgroups of returnees and HSs against each other to establish to what extent any lower performance of the returnees is due to attrition.

RQ3: Is there an effect of L1 Arabic on knowledge and processing of L2 English collocations?

We expect to find little evidence for CLI among HS groups, because they had had relatively little Arabic contact in the US, while the strongest impact of CLI was expected among the child returnees. It was predicted that returnees would underperform compared with the HSs across all collocation conditions. The returnees were also expected to respond more quickly to congruent collocations than to incongruent ones.

Methods

Participants

A quasi-experimental, cross-sectional design was chosen. Participants were allocated to different groups based on their language learning history since random allocation was not possible. A total of 118 Arabic-English bilinguals allocated to five groups participated in this study, namely 23 child returnees, 21 adolescent returnees, 26 child HSs, and 28 adolescent HSs. Finally, 20 adult HSs functioned as a base line. Demographic information for each group is presented below.

The study involved 44 child and adolescent returnees aged between 20 and 45 years (mean age 31.45), who grew up bilingually in the US as second-generation migrants and had returned to their homeland, SA, at different points in time. All returnees were born in the US or had moved there before the age of five. The primary criterion which distinguishes the returnees is the AoR which ranges from age five to seventeen. This variable allows their division into two main subgroups: 23 participants who returned to SA up to the

[1]*It is noteworthy that all returnees in this study attended public (state) schools in SA, supporting the expectation of Arabic dominance among the participants. However, it is important to consider that this may not always be the case. Sometimes young returnees enrol in international schools and are immersed in an English-speaking environment, studying and communicating primarily in English. This exposure can influence their language dominance, potentially leading them to be more dominant in English than Arabic. However, this was not the case for the current sample.*

https://doi.org/10.1017/S1366728924000610 Published online by Cambridge University Press

age of 11, referred to as, child returnees (RT1) represent the pre-puberty stage, and 21 participants who returned at or after the age of 12 to SA, referred to as, adolescent returnees (RT2), represent the post-puberty stage. The decision to set the cut-off point at the age of 12 has been made based on previous literature that considers that there is a change in attrition susceptibility at around age 12 (Bylund, 2019). It is important to consider the research on sensitive periods for lexis and collocational abilities. Although research on this topic is scarce, available findings suggest that acquisition within this domain is also subject to maturational constraints, indicating that collocational abilities have a peak period of sensitivity ranging from 0 to 6 years, followed by an offset period lasting between 6 and 12 years, possibly around age 9 (Granena & Long, 2013).

The returnees' LoR in SA ranges from 11–38 years. The length of stay in SA, also known as the 'incubation period' or 'length of attrition,' refers to the time elapsed between the participants' return to SA and the first test session. This is seen as an important factor in language attrition and many authors establish minimal baselines after which attrition effects may occur. However, researchers have not yet confirmed that the incubation period, despite being intuitively crucial, is indeed a cause for a language to attrite (Mehotcheva & Köpke, 2019). The exact point when someone is likely to become incapable of speaking a previously fluent L2 is unclear (Larson-Hall, 2019). A minimum of a ten-year stay in the new linguistic environment was taken as one of the inclusion criteria since it is a widely accepted baseline in attrition literature (Gürel, 2004). A well-known study of Spanish L2 attrition is found Graham (2012). After twelve years of incubation, Graham studied participants who spent twenty-four months abroad and found that they lost a significant number of tokens on a narrative task, but they still managed to function in their L2 Spanish. In light of that, for RT1 a minimal length of stay of 11 years in SA was specified, and for RT2, a minimum of 12 years. On the basis of the available literature, it was assumed that L2 attrition might be detected after these periods of time had elapsed.

The returnees were compared to 54 US-based child and adolescent Saudi HSs of Arabic aged from six to seventeen years old (mean age 11.85) and finally a group of 20 AHS aged from nineteen to thirty that functions as a base line. The HSs and returnees are assumed to be comparable (Treffers-Daller et al., 2016) because the returnees belong to the same group of the HSs up until the time where the returnees moved back to SA. Similar to the division of the returnee subgroups, there were 26 child HSs (HS1) aged between 6–11, and their 28 adolescent counterparts (HS2) aged between 12–17. Figure 1 illustrates the division into groups.

All participants had Arab parents from SA and had acquired Arabic as their L1 and English as L2. The majority came from middle class or upper middle-class backgrounds. The parents had either a bachelor or a post-graduate degree. Participants were recruited through a snowball sampling method since it was not possible to randomly sample informants (see Table 1 for further details).

Vocabulary Tasks

In order to determine the participants' vocabulary knowledge of English, the Peabody Picture Vocabulary Task (PPVT) (Dunn & Dunn, 2007) was administered. This task has been widely used to measure the receptive vocabulary knowledge among children and

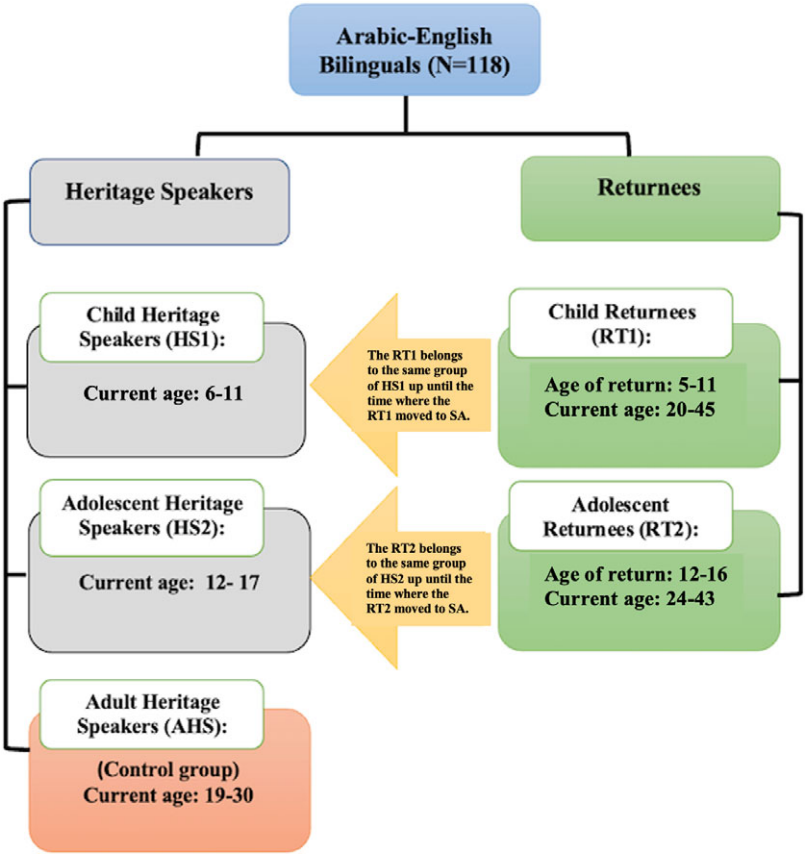

Figure 1. The HS and RT groups in the current study.

https://doi.org/10.1017/S1366728924000610 Published online by Cambridge University Press

Table 1. Overview of Participants

Groups		Arabic-English Returnees		Arabic-English HSs		
Subgroups		Child Returnees	Adolescent Returnees	Child HSs	Adolescent HSs	Adult HSs
Group Code		RT1	RT2	HS1	HS2	AHS
Place of Residence		SA	SA	US	US	US
N		23	21	26	28	20
L2 Onset	Mean (SD)	1.34 (1.83)	2.23 (1.87)	1.84 (1.87)	2.75 (1.93)	1.45 (1.94)
	Range	0–5	0–5	0–5	0–5	0–5
Age of Return	Mean (SD)	7.69 (1.76)	13.43 (1.26)	N.A.	N.A.	N.A.
	Range	5–11	12–16			
Age at Recording	Mean (SD)	28.9 (6.07)	34.23 (5.35)	9.11 (1.52)	14.39 (1.73)	22.4 (3.26)
	Range	20–45	24–43	6–11	12–17	19–30
LoR in US (years in the US)	Mean (SD)	7.87 (2.11)	13.42 (1.25)	Since birth	Since birth	Since birth
	Range	0–11	12–16			
LoR in SA (incubation period/years in SA)	Mean (SD)	21.04 (6.45)	20.80 (5.55)	N.A.	N.A.	N.A.
	Range	11–38	12–31			

adult bilinguals. It has also been used in several studies on L2 attrition (e.g., Tomiyama, 2009).

Since participants' Arabic knowledge might differ, an Arabic vocabulary size test referred to as Arabic-Lex (Masrai & Milton, 2019) was used. The aim of this test was to assess the Arabic speakers' written receptive vocabulary knowledge of the 50,000 most frequent Arabic words. It comprises 120 test items, including 20 non-words which were inserted randomly throughout the test. An adult version and a child version of the test were available (see Appendix S1 and S2 in Supplementary Material).

Background Questionnaires

A questionnaire was adapted from the Bilingual Language Profile (BLP) (Birdsong et al., 2012) to assess bilinguals' language dominance. The highest achievable score for one language is 218, indicating a high level of proficiency, and a significant exposure to and motivation for the target language. Subtracting the total score of one target language from the other yields the dominance index. The global dominance score ranges from −218 to +218; a negative score indicates Arabic dominance, whereas a positive score indicates English dominance. A score close to 0 implies similar results for both languages, indicating that the individual is likely to be a balanced bilingual. MacArthur's subjective social status scale was employed to assess participants' social status.

Digit Span Tasks

The backward digit span task (DST) was administered as a measure of working memory which is part of the Wechsler IV Adult Intelligence Scale (Wechsler, 2008). A backward DST was chosen rather than a forward DST because the former measures complex verbal working memory capacity and is strongly associated with academic ability and cognitive performance, whereas the latter imposes a minimal processing load and only measures short-term memory. Since bilinguals constantly activate both languages in language processing, the task was administered in both English and Arabic.

The Selection of Collocations

To ascertain which of the selected English collocations had an Arabic translation equivalent (congruent/Arabic-English) and which ones did not (incongruent/English-only), an Arabic-English bilingual dictionary and four Arabic NSs were consulted. It is noteworthy that word order inside the collocations is the same in Arabic and English, as they are all VN, and (in)congruence therefore relates to the existence of a literal translation equivalent between the two languages. Other non-existing English collocations with (Arabic-only) and without Arabic equivalents (baseline) were added. Since semantic transparency plays an important role in collocational processing (Gyllstad & Wolter, 2016), three English NSs were consulted to check if the Arabic-only collocations that were created by translating Arabic collocations into English were semantically transparent. They confirmed that this was indeed the case, as they were able to explain the meaning of the novel collocations. The NSs were also asked to complete the tasks before giving them to participants and they achieved high scores. They were then asked to judge the tasks based on difficulty and clarity. Based on their feedback, some collocations were excluded due to their difficulty, resulting in a total number of 92 VN collocations.

The items were then classified into four categories: (1) congruent collocations (Arabic-English), (2) English-only (incongruent) collocations/non-existing in Arabic, (3) Arabic-only (translated) collocations/non-existing in English and (4) baseline items that are non-existent either in English or Arabic. The words were recombined from the other three categories to create the baseline items. It was done to ensure the lexical frequency of individual words was kept constant across different conditions (see Wolter & Yamashita, 2015). Each category consists of an equal number of 23 collocations (see Appendix S3 in Supplementary Material for the complete list). Table 2 shows an example of collocation categories used in the study.

https://doi.org/10.1017/S1366728924000610 Published online by Cambridge University Press

Table 2. Example of collocation categories

Arabic-English	English-only	Arabic-only	Baseline
achieve success	make mistakes	*eat money	*do effort

A frequency-based approach was used to identify these collocations. Several items were chosen from the phrasal expression list that contains the most frequent English MWUs derived from the British National Corpus (BNC), which was compiled by Martinez and Schmitt (2012), such as *take advantage* and *make sense*. Martinez and Schmitt's primary criteria for selection was to include items that are identified to pose difficulties for English learners, particularly at a receptive level. Only two-word collocations were chosen to avoid variability in results due to differences in collocation length. To further examine the English collocations, the corpus of contemporary American English (COCA; Davies, 2008) was used as a reference corpus because the HSs lived in the US at the time of testing and the returnees had studied in the US before their return to SA.

We used Nguyen and Webb's (2017) criterium for selecting collocations: all English-Arabic and English-only collocations had a frequency of at least 50 in the COCA, with a minimum Mutual Information (MI) score of 3, which indicates a substantial collocational link (Hunston, 2022). An Arabic corpus (arTenTen24) on Sketch Engine was used to ensure appropriate categorisation. Moreover, both corpora were used to verify that translation equivalents of Arabic-only and English-only collocations did not exist in the other language. While some Arabic-only and baseline collocations registered a small number of occurrences, they showed a negative MI score which indicates dissociation rather than association between the two words, instead of significant co-occurrence in English (Wolter & Yamashita, 2018). Accordingly, the information from COCA indicated that the categorisation of the items was appropriate.

Gap-filling Task

A Gap-filling task (GFT) was employed to investigate participants' ability to produce English collocations as well as their accuracy and performance speed. The task included the same VN collocations except for the non-existing English collocations. The experiment was designed in PsychoPy, an open-source experimental software for running online cognitive experiments which taps into processing (Peirce, 2007). All sentences were extracted from the BNC and presented in random order. They had a minimum of 95% lexical coverage which has been suggested as a reasonable threshold for reading comprehension (Laufer, 1989). In each sentence, participants were asked to fill in the blanks by typing the missing verb as quickly and correctly as possible. Sentences appeared one at a time in the middle of the screen with one blank. The first letter of the missing collocate was provided to restrict variability in participants' answer options. Spelling errors (e.g., **breik* instead of *break*) or incorrect verb forms [e.g., **maked* instead of *made*) were not considered in the analysis if the completed word was lexically correct (Nesselhauf, 2003). To control for the effects of sentence length on participants' responses, the blank was placed on the fifth word across all sentences and sentence length was kept consistent. Due to children's slower typing speed, the researcher read out the sentences to them and concurrently typed on their behalf. However, considering the potential for online connection disruptions during COVID-19, alongside variations in the researcher's articulation time and typing speed, the reaction time data for this task was excluded from the analysis.

Acceptability Judgement Task

The AJT included all 92 collocations in random order, designed in PsychoPy. Participants were required to judge whether an English collocation was an existing collocation or not by pressing one of the two keys 'a' and 'k', which represented yes and no, respectively. They were informed that they should answer as quickly and accurately as possible. Prior to the experimental session, participants were presented with instructions and were given a practice session to familiarise themselves with the task. A fixation point then appeared in the middle of the screen for 500 milliseconds, followed by a test item that remained on the screen until the participant responded. A limit of 7000 milliseconds was chosen as timeout.

Procedure

The tasks were counterbalanced across participants in that half of the participants from each subgroup completed the English tests first and the other half started with the Arabic tests. Participants completed the tasks in English and Arabic on different days, to avoid participants being in a bilingual mode as much as possible, because this could have led to increased CLI. All instruments were administered online via Pavlovia, as face-to-face data collection was not possible during COVID-19. The tasks lasted approximately one hour in total for each participant.

Data Analysis

Generalized linear mixed effect modelling was used for accuracy and linear mixed effect modelling for reaction time. For both the AJT and the GFT, a model was constructed with accuracy as dependent variables. Reaction time was only analysed for the AJT. Group, condition, and interaction between group and condition were included as fixed effects. Length was also included as a fixed effect to adjust for character length in the reaction time models. Participant and item were included as random intercepts to capture the variability and the individual differences. Random slopes by-participant for condition were added to explore how the impact of condition differs across participants. Sum-coding was employed for the independent variables, specifically for group and condition. This meant setting one level as negative and another as positive, with zero as the mean, resulting in the contrast vector (−1,1). Variables were scaled to bring all variables on a similar scale ensuring that no single variable dominates the analysis due to its larger magnitude. After thoroughly evaluating multiple models and applying the forward method approach, the best-fitting model was identified (Barr et al., 2013). The models were fit with R (R Core Team, 2013) version 2023.12.1 + 402 with the package lme4 (Bates et al., 2015). For collinearity issues, the variance inflation factor (*VIF*) function from the *Car* package was used. All *VIF* values were below the threshold of 10, confirming that there were no issues with multicollinearity (Jou et al., 2014). One-way ANOVAs were conducted to further examine between group differences.

Prior to conducting the AJT analyses with the reaction times as the dependent variable, all inaccurate trials and trials that were accurate but took less than 200 ms were removed. Inaccurate trials included answering "no" to real-word items and items which participants failed to answer because they ran out of time. Reaction times were log-transformed as a correction for non-normality. Outliers were not removed since the data had undergone log transformation. According to Nicklin and Plonsky (2020), using a log-transformation

https://doi.org/10.1017/S1366728924000610 Published online by Cambridge University Press

is an effective way to deal with reaction time outliers. As such, it has been shown to effectively reduce the influence of slow-response outliers while maintaining statistical power.

Results

English and Arabic Vocabulary Tasks

A one-way ANOVA was conducted to test whether the differences between groups on the PPVT were statistically significant. Results showed a significant difference (ANOVA, F (4,113) = 42.14, $p = 0.001$) with the HS1 obtaining the lowest score. Post-hoc results revealed no significant differences between the RT2, the HS2 and the AHS, indicating that they performed similarly. However, the RT1 scored significantly lower than all the others. There was also a significant difference on the Arabic task (ANOVA, F (4,113) = 41.81, $p = 0.001$) with the AHS obtaining the lowest score, while the RT1 obtaining the highest score. Results of the post-hoc showed no significant differences between the RT1 and the RT2, implying that they scored nearly similar results and demonstrated statistically significant higher scores compared to the HSs.

Vocabulary dominance was computed based on both the English and Arabic vocabulary tasks. Figure 2 shows the subtraction-derived dominance indices among the five bilingual groups. According to the between-language subtractive differential plot, the HS2 and the AHS are clearly English-dominant, while the RT1 are Arabic-dominant. However, the RT2 showed a nearly zero between-language subtractive differential, indicating high balance. The HSs scored lower on the Arabic vocabulary task than on the English one. Their preference for English is evident in the positive between-language proficiency differential. Nevertheless, a large variability in scores indicates significant diversity in vocabulary dominance. As expected, the RT1 scored lower in English than in Arabic. The negative differential indicates Arabic dominance among the RT1.

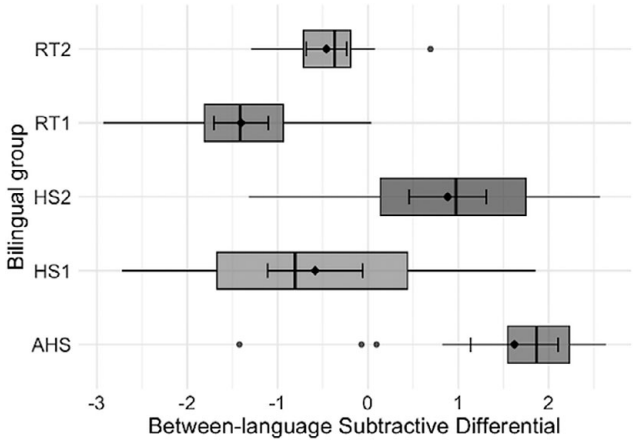

Figure 2. Vocabulary Dominance indices as a function of participant group, based on the English and Arabic Vocabulary tasks calculated by the differential method (values close to 0 indicate balanced dominance, negative values for dominance towards Arabic, positive values for dominance towards English).

Language dominance

Figure 3 plots subtraction-derived dominance indices as measured by the BLP questionnaire among all groups. The between-language subtractive differential plot showed that all HSs were clearly English-dominant, whereas the RT1 were Arabic dominant. The RT2, however, showed a near-zero between language subtractive differential, indicating a high degree of balance. The HSs scored lower on Arabic usage than on English usage. The positive between-language proficiency differential manifests their preference for English. The minimum and maximum values obtained through differential score also indicate that no participant in this group was Arabic-dominant (no negative scores). As expected, the RT1 obtained higher scores in Arabic than English. The negative differential indicates Arabic dominance among RT1.

As the BLP data was normally distributed, a one-way ANOVA was conducted to examine variability in the extralinguistic variables measured by the BLP questionnaire, revealing significant differences between groups. Post-hoc results showed significant differences between groups in terms of their Arabic use in which the the RT1 scored significantly higher than the others, indicating a higher level of Arabic usage among the RT1. As for English use, post-hoc results indicated that the returnees scored significantly lower than

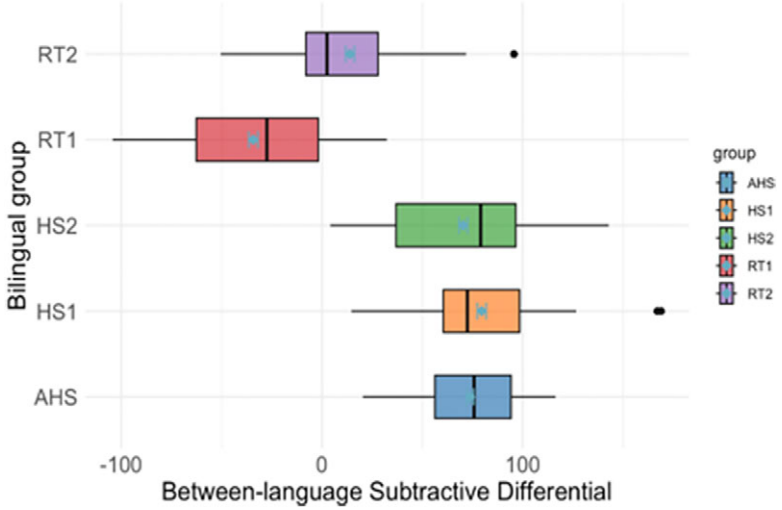

Figure 3. Language Dominance indices as a function of participant group, based on the BLP calculated by the differential method (values close to 0 indicate balanced dominance, negative values for dominance towards Arabic, positive values for dominance towards English).

https://doi.org/10.1017/S1366728924000610 Published online by Cambridge University Press

the HSs and that the returnees had demonstrated notably lower levels of English usage.

Digit Span Task

A one-way ANOVA test was used to determine whether the differences of the English and Arabic Backward DST across groups were statistically significant. As for the English DST, results showed a significant difference between groups (ANOVA, F (4,113) = 15.67, p = 0.001) with the RT1 obtaining the lowest score. There was also a significant difference between groups on the Arabic DST (ANOVA, F (4,113) = 20.56, p = 0.001), with the HS1 obtaining the lowest score. Post hoc show that returnees scored significantly higher on the Arabic DST than the HSs.

Table 3. Accuracy results for the Gap-filling Task

Accuracy- Gap-filling Task - Formula: accuracy ~ group * Condition + (1 | participant) + (1 | item)				
Fixed effects	Estimate	SE	*z*	*p*
(Intercept)	4.84	0.53	9.11	< 2e–16 ***
groupHS1	−1.33	0.42	−3.16	0.001 *
groupHS2	−0.21	0.44	−0.48	0.63
groupRT1	−2.44	0.42	−5.79	6.91e–09 ***
groupRT2	−0.92	0.44	−2.06	0.03 *
English-only	0.12	0.55	0.22	0.82
HS1: English-only	−0.67	0.23	−2.98	0.002 **
HS2: English-only	−0.22	0.28	−0.78	0.43
RT1: English-only	0.13	0.23	0.57	0.56
RT2: English-only	0.43	0.30	1.44	0.15

Gap-filling task

The summary of the models for accuracy for the GFT is presented in Table 3. There was a significant effect of group in that the AHS and the HS2 had significantly higher scores than the others (see Figure 4). There was no significant difference between the AHS and HS2 (E = −0.21, z = −0.48, p = 0.63), implying that they exhibited similar scores. Post hoc estimated marginal means (EMMeans) with Bonferroni correction comparisons showed no significant differences between the RT2 and their counterpart HS2 (E = 0.37, z = 1.03, p = 1.000), and no significant differences between the RT1 and their counterpart HS1 (E = 0.71, z = 2.24, p = 0.24). There was an interaction between group and condition only for the HS1 group (E = −2.98, z = −0.68, p = .002), indicating that they performed significantly better on congruent trials than on incongruent ones. However, no interactions between group and condition were observed among returnees, suggesting that they were equally good at the different conditions. Therefore, there was little evidence for Arabic influence on collocational knowledge among returnees.

Acceptability Judgement Task

Tables 4 and 5 present the summary of the models for both accuracy and reaction time for the AJT. There was a main effect of group for

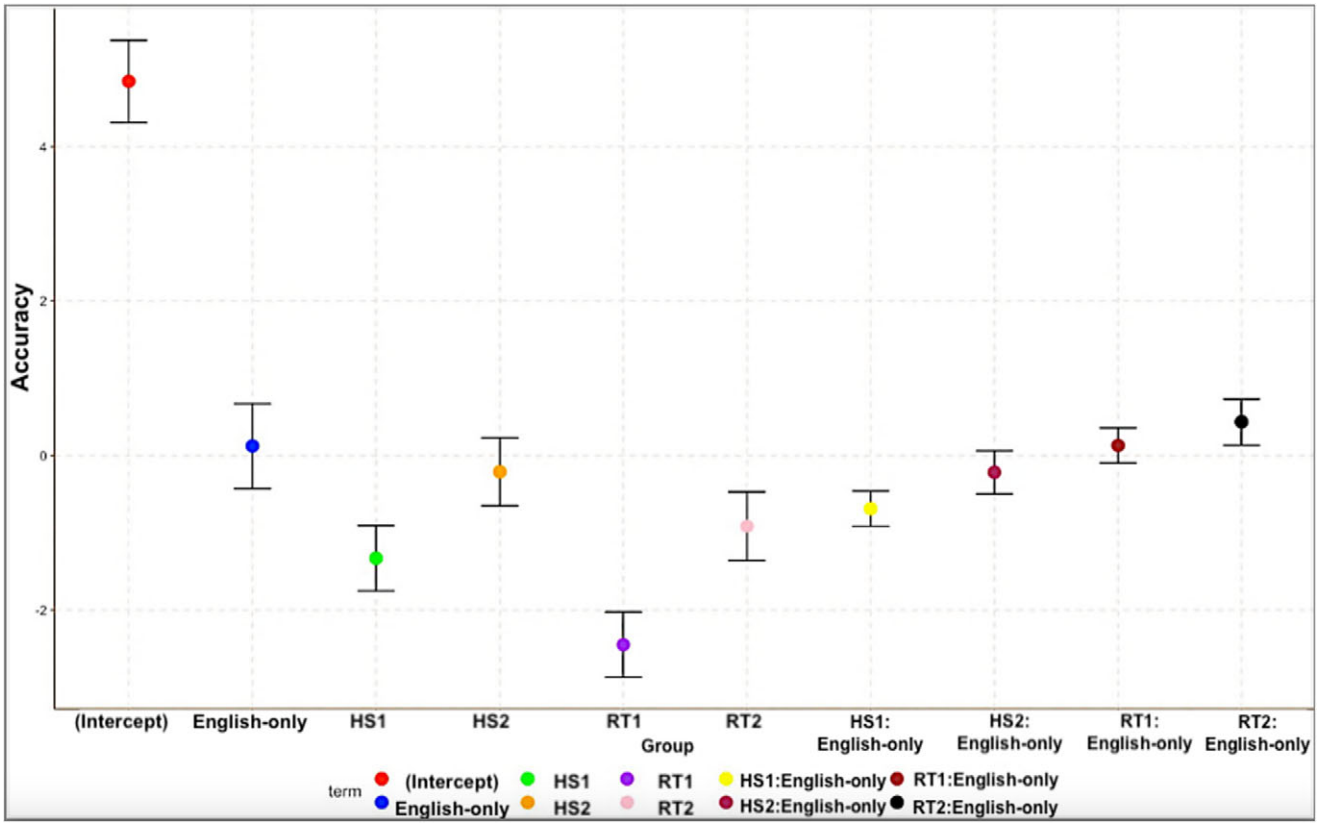

Figure 4. Estimated Coefficients of Accuracy for the Gap-filling Task with Standard Error Bars.

https://doi.org/10.1017/S1366728924000610 Published online by Cambridge University Press

Table 4. Accuracy results for the AJT

Accuracy – AJT - Formula: accuracy ~ group * condition + (1 | participant) + (1 | item)				
Fixed effects	Estimate	SE	*z*	*p*
(Intercept)	3.47	0.24	14.33	< 2e–16 ***
groupHS1	−0.51	0.19	−2.60	0.009 **
groupHS2	0.07	0.21	0.33	0.74
groupRT1	−0.41	0.21	−1.97	0.04 *
groupRT2	−0.12	0.23	−0.51	0.61
English-only	0.28	0.34	0.83	0.40
Arabic-only	−2.32	0.32	−7.36	1.82e–13 ***
Baseline	−1.68	0.32	−5.29	1.21e–07 ***
HS1: English-only	−0.32	0.24	−1.31	0.19
HS2: English-only	0.03	0.27	0.100	0.92
RT1: English-only	−0.12	0.26	−0.45	0.65
RT2: English-only	0.35	0.31	1.12	0.26
HS1: Arabic-only	1.04	0.19	5.33	9.94e–08 ***
HS2: Arabic-only	0.36	0.21	1.72	0.09
RT1: Arabic-only	−0.43	0.19	−2.16	0.03 *
RT2: Arabic-only	−0.38	0.22	−1.75	0.08
HS1: Baseline	0.56	0.19	2.80	0.005 **
HS2: Baseline	0.17	0.22	0.76	0.44
RT1: Baseline	−0.36	0.20	−1.75	0.08
RT2: Baseline	−0.39	0.23	−1.73	0.08

Table 5: Reaction time results for the AJT

Reaction time-AJT: Formula: log_rt_AJT ~ length + group * condition + (1 | participant) + (1 | item) + (1 + condition | participant)				
Fixed effects	Estimate	SE	*t*	*p*
(Intercept)	−0.16	0.09	−1.77	0.08
groupHS1	0.74	0.11	6.79	5.37e–10 ***
groupHS2	0.43	0.11	4.07	8.60e–05 ***
groupRT1	0.30	0.11	2.74	0.007 **
groupRT2	0.09	0.11	0.82	0.42
length	0.02	0.003	6.76	1.65e–09 ***
English-only	0.001	0.03	−0.02	0.97
Arabic-only	0.24	0.04	5.82	3.32e–08 ***
Baseline	0.19	0.03	5.05	1.29e–06 ***
HS1: English-only	−0.03	0.04	−0.71	0.48
HS2: English-only	−0.01	0.04	−0.32	0.74
RT1: English-only	−0.08	0.04	−2.03	0.04 *
RT2: English-only	−0.07	0.04	−1.73	0.08
HS1: Arabic-only	−0.15	0.05	−3.20	0.001 **
HS2: Arabic-only	−0.05	0.05	−1.02	0.31
RT1: Arabic-only	0.21	0.05	4.07	8.36e–05 ***
RT2: Arabic-only	0.17	0.05	3.26	0.001 **
HS1: Baseline	−0.12	0.04	−2.72	0.007 **
HS2: Baseline	−0.03	0.04	−0.71	0.48
RT1: Baseline	0.05	0.05	0.99	0.32
RT2: Baseline	0.07	0.05	1.59	0.11

accuracy in that the AHS, the HS2, and the RT2 had significantly higher scores than the others. However, post-hoc EMMeans results revealed a significant difference between RT2 and their counterpart HS2 ($E = 0.87$, $z = 4.45$, $p = 0.001$), which shows that HS2 obtained higher scores than RT2. A significant difference was also observed between RT1 and their counterpart, HS1 ($E = 1.09$, $z = 4.51$, $p = 0.001$), indicating that HS1 achieved higher scores than RT1. Figure 5 illustrates that there is considerable variability within the younger HSs groups, yet they performed better compared to the returnee groups. There was an interaction between group and condition for the HS1 in that they performed significantly less well on Arabic-only ($E = 1.04$, $z = 5.33$, $p < 0.001$) and baseline collocations ($E = 0.56$, $z = 2.80$, $p = 0.005$). An interaction was also found between RT1 and condition in that the RT1 performed significantly less well only at Arabic-only collocations ($E = -0.43$, $z = -2.16$, $p = 0.03$).

As for RT, there was a main effect of group in that the AHS and the RT2 were significantly faster than the others. Post-hoc EMMeans results revealed no significant difference between the RT2 and their counterpart the HS2 ($E = 0.27$, $z = 2.64$, $p = .08$), implying that they exhibited similar scores. Conversely, a significant difference was seen between the RT1 and their counterpart, the HS1($E = 0.31$, $z = 3.03$, $p = .02$), indicating that the RT1 were faster than HS1. There was also a main effect for length of collocations ($E = 0.033$, $t = 2.71$, $p = .007$), suggesting that as word length increased, participants tended to take longer time to respond. A main effect of condition was observed in that the non-existing collocations were recognised more slowly than the existing congruent collocations. An interaction was observed between group and condition for the RT1 ($E = 0.21$, $z = 4.07$, $p < .001$) and RT2 only ($E = 0.17$, $z = 3.26$, $p = .001$) in that they were slower at Arabic-only collocations. Furthermore, there was an interaction between group and condition in which the HS1 responded faster at Arabic-only ($E = -0.15$, $z = -3.20$, $p = .001$) and baseline collocations ($E = -0.12$, $z = -2.72$, $p = .007$). An interaction was also seen between group and condition, indicating that RT1 responded faster at incongruent collocations ($E = -0.08$, $z = -2.03$, $p = .04$). A one-way ANOVA revealed that the RT2 performed significantly faster on incongruent trials than on congruent trials (ANOVA, $F(1,19) = 29.03$, $p < 0.05$).

Summary of the findings

In terms of language dominance, the results showed, first of all, that all HSs were clearly English-dominant, whereas the RT1 were Arabic-dominant. However, the RT2 were the most balanced group. Regarding accuracy on the GFT, AHS and the HS2 scored significantly higher than the others. Interestingly, no significant differences were found between the RT2 and their counterpart HS2, nor between RT1 and their counterpart HS1. Therefore, there was little evidence for Arabic influence on the productive task. As for accuracy on the receptive task, the AHS, the HS2, and the RT2 had significantly higher scores than the others. However, HS1 and HS2 significantly outperformed their counterparts RT1 and RT2. RT1 performed significantly less well only on Arabic-only collocations.

https://doi.org/10.1017/S1366728924000610 Published online by Cambridge University Press

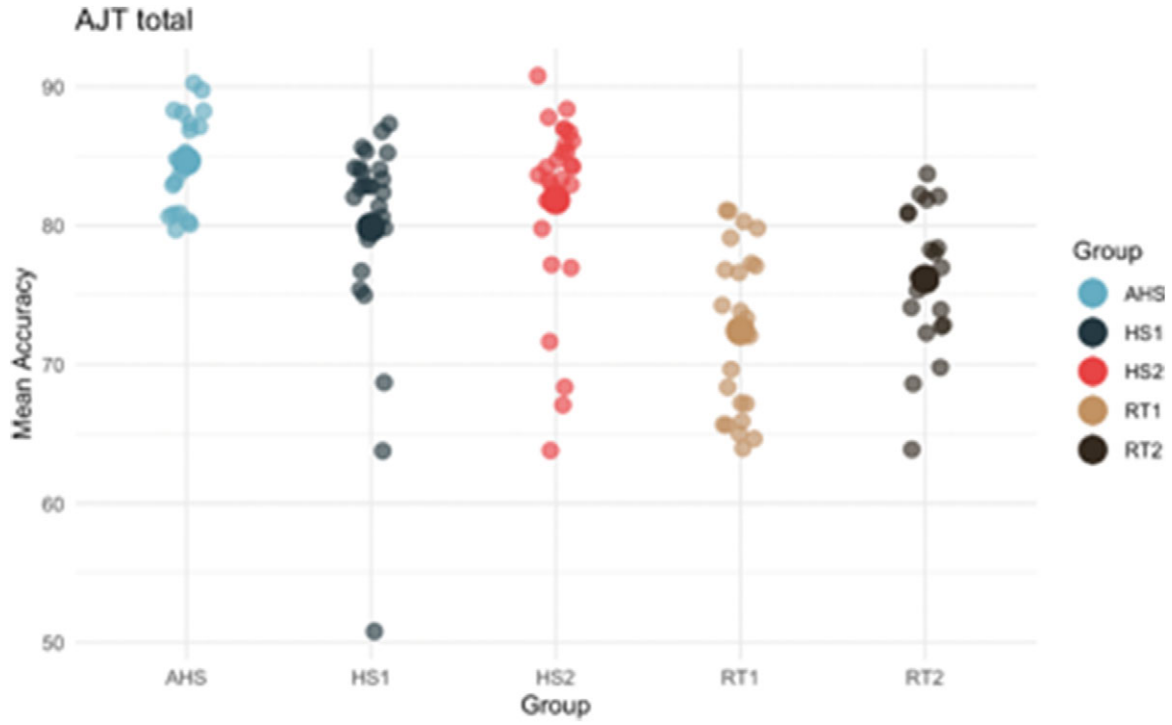

Figure 5. Total mean Accuracy results for the AJT.

Regarding reaction times on the AJT, AHS and RT2 were significantly faster than the others. Returnees were slower at Arabic-only collocations, and they responded significantly faster for incongruent collocations than for congruent ones.

Discussion

RQ1 aimed at understanding whether the groups differ from each other with respect to language dominance. It is evident from the comparison between the five groups that this is indeed the case. The results revealed that language dominance differed considerably by group, both with respect to general dominance as measured with the BLP and dominance at the level of English and Arabic vocabulary. The HSs living in the US were clearly L2 dominant. The results are in line with prior research about HL development, which has shown the L2 becomes dominant once HSs enrol in the L2 school system (Kupisch et al., 2021). Our study shows that L2 dominance can also be observed with AHS. By contrast, as predicted, the RT1 were Arabic dominant. These returnees left the US between ages five and eleven yet exhibited strong Arabic usage either within the family or at work. In this sense, AoR plays a crucial role in the process of dominance shift. This outcome confirms findings that suggest that balanced bilingualism is unlikely to happen if return to the homeland happens in early childhood (Flores et al., 2022).

Regarding the RT2, we predicted that they would be the most balanced group because of their extensive exposure to both languages due to their extended stay in the US prior to moving back to SA. Interestingly, the RT2 were the most balanced group on both the computations of language dominance for the BLP and the vocabulary tasks. That is, in the Arabic test, the performance of the RT2 was similar to that of the RT1 who obtained the highest score across groups. In the same way, the RT2 performed on a par with the AHS, who achieved the highest score on the PPVT. This finding is consistent with other studies that have found that longer periods of exposure were linked to better scores among returnees (Flores, 2010). However, RT2 scored significantly lower on Arabic usage compared to RT1. This outcome also supports what Dörnyei et al. (2004) describe as "immersion" and "acculturation" as central modifying factors that facilitate the overall process of language learning. As discussed, it may not only be exposure itself that holds importance, but rather the quality of engagement with the language that takes place in a socially integrated environment. Thus, it is the amount of contact with both languages, prior to and after return, but not just LoR in the US that explains the bilingual's performance. This outcome confirms previous findings of Matos and Flores (2022) who suggest that bilinguals' language competence is not affected by reduced exposure per se, but by their type of high-quality engagement with language during this time. Several studies, focusing primarily on L1 attrition in migration contexts (rather than returnees), have demonstrated that attrition effects in the lexical domain are not only determined by the amount of contact, but also by the type of contact (e.g., professional contexts) (Schmid & Jarvis, 2014).

RQ2 sought to determine whether there is evidence for attrition in the productive and receptive English knowledge collocations among returnees. To begin with the GFT, results showed that, as predicted, the AHS outperformed the returnees in accuracy. The accuracy data also revealed that the HS1 performed less well than HS2, which seems to indicate that the HS1 had not yet acquired these collocations, suggesting incomplete acquisition for the HS1 only, while the HS2, who had had more contact with English prior to return did not differ in performance from the AHS. This finding is consistent with Bylund's (2019) proposal that pre-puberty immigrant children may not have the same levels of linguistic knowledge as post puberty immigrants, whose performance may be within the range of that of monolinguals on various tasks. Furthermore, since the HS2 obtained scores similar to the AHS scores, we can assume that the RT2 (who left the US at ages similar to the HS2) had

https://doi.org/10.1017/S1366728924000610 Published online by Cambridge University Press

acquired the collocations prior to return. Thus, the RT2 might indeed be in an attrition scenario. However, post hoc results comparing subgroups failed to reveal significant differences between the RT2 and its counterpart HS2, and between the RT1 and its counterpart HS1, possibly due to the lack of statistical power for comparisons between subgroups.

The AJT results showed a different picture. Contrary to expectations, the younger HS groups (HS1 and HS2) achieved significantly higher scores in accuracy than the corresponding returnee groups (RT1 and RT2). Although the returnees are adults, they performed significantly less well than their HS counterparts. This indicates that the younger HSs had already acquired these collocations and were familiar with them at the time of data collection. Conversely, this makes it more likely that the returnees' poor performance on this task is the result of attrition, at least for receptive tasks. Clearly, this shows that pattern recognition skills, which includes the ability to recognize collocations, may diminish under conditions of reduced input (Arnon et al., 2017). The contribution of this study is that it has demonstrated such skills can indeed attrite among returnees, under conditions such as those experienced by child returnees.

On the other hand, as for reaction times, the AHS and the RT2 were significantly faster in responding than the others. This finding confirms our hypothesis that the RT1 had not yet acquired these collocations before returning. Upon comparing the groups between each other, results showed that the RT2 were as fast as the HS2, but the RT1 significantly outperformed their counterpart, HS1. This could be because adults are quicker and more experienced with handling computers than children. Another possible explanation could be that the ability to recognize existing L2 collocations depends upon the amount of contact with the L2 after return, regardless of whether their return is early or late, as has been demonstrated for morphology by Matos and Flores (2022). If so, the evidence presented supports emergentist theories of language acquisition. The most prominent ones are usage-based models that assert that language experience is a key predictor of linguistic knowledge and it is therefore likely that extensive exposure to a language reflects a higher self-reported proficiency level (Bybee, 2006). The fact that the RT2 group performed on a par with the reference group, the AHS and the HS2, may be explained by the fact the RT2 group were balanced bilinguals who used English frequently on a daily basis, contrary to the RT1 group, who were clearly Arabic-dominant.

The differences between the results of the receptive and productive tasks need to be discussed in more depth. We assume that participants who had lost L2 input throughout their adolescent years (the RT2) did not have problems producing the collocations on the GFT, but found it difficult to recognize them correctly on the AJT which was somewhat unexpected. These findings are inconsistent with existing research (e.g., Alharthi, 2015; González-fernández & Schmitt, 2020; Nesselhauf, 2003) emphasizing that the development of productive vocabulary involves complex cognitive processes and bilinguals experience greater difficulties on collocations in productive tasks compared to receptive ones. In this study, the observed difficulty in the receptive task has posed greater challenges to returnees when compared to the productive task. A possible explanation could be that their L1 may have been more activated during the receptive task, because of the presence of Arabic-English and Arabic-only collocations. In contrast, Arabic was perhaps less activated in the GFT because there were no Arabic-only collocations in this task. That L1 activation might explain the differences between both tasks is consistent with the predictions regarding linguistic accessibility in bilinguals known as the Activation Threshold Hypothesis (ATH) (Paradis, 2007). Bilinguals generally have difficulty finding words since they need to inhibit the language that is not being activated (Bialystok et al., 2012), in this case Arabic. However, it is also possible that some bilingual returnees, particularly the child returnees, find it more difficult to access and retrieve lexical items from English because they no longer use the language on a daily basis. Instead, the constant use of L1 hinders the activation of L2 on the receptive task.

Furthermore, the RT1 performed significantly less well than the RT2 on both tasks, and it was the only group that exhibited an L1 effect on Arabic-only collocations. This finding indicates that the AoR is an important variable explaining, at least in part, the returnees' performance on tasks. Thus, the current study lends some support to the assumption that a stabilization period is needed, also for the acquisition of collocational knowledge. In other words, the earlier the returnees moved back, the greater the likelihood of L2 attrition. This is consistent with previous literature on returnees (Flores, 2010, 2019; Flores et al., 2022), suggesting that attrition effects in returnees emerge immediately after return, at least for returnees who move to their homeland during childhood. However, when the return occurs during adolescence, signs of attrition are more difficult to detect. Thus, the present study reveals that the younger the child is upon return, the more pronounced signs of attrition become.

RQ3 focused on whether there is an effect of L1 Arabic on knowledge and processing of L2 English collocations. Unexpectedly, no strong evidence of Arabic influence was observed among returnees on the GFT. As for reaction times on the AJT, one unanticipated finding was the interaction between group and condition for RT1 only, revealing faster responses to incongruent trials. Moreover, the RT2 demonstrated significantly faster response times to incongruent trials and recognized congruent trials significantly more slowly, which contradicts evidence from previous research (Yamashita & Jiang, 2010) that found that L2 learners acquire congruent L2 collocations quicker and more accurately than incongruent collocations. One possible explanation for this might be that participants were slowed down by the congruence between languages, which might have led to increased activation of the Arabic translation equivalent. Suppressing this translation equivalent might be costly, resulting in increased reaction times. The results might also be explained by what Kupisch (2014) refers to as CLO; that is, the RT2 tended to 'over-inhibit' the structure that is similar in both languages, in an attempt to avoid influence of the societally dominant language, in this case Arabic, while over-emphasizing the differences with English. Therefore, they struggled to produce English collocations with an Arabic equivalent correctly. Their slower response towards congruent trials may tentatively be attributed to their awareness of the similarities and differences between English and Arabic, which may have led to hesitation and attempts to avoid any potential errors from Arabic transfer. However, they might not exhibit the same uncertainty with English collocations without an Arabic equivalent since crosslinguistic influence is less likely for these collocations.

Regarding the AJT, as expected, there was L1 influence on the processing of L2 collocations among returnees. A significant interaction was found between group and condition indicating that the RT1 performed less well only at the condition of non-existing English collocations with an Arabic equivalent in accuracy and reaction times. This outcome is in line with previous research (e.g., Flores, 2010) that has found evidence for CLI from L1. A possible explanation for this could be that the word order of VN collocations is the same in English and Arabic, whereas they often mismatch in adjective-noun collocations. As noted by

https://doi.org/10.1017/S1366728924000610 Published online by Cambridge University Press

Müller and Hulk (2001), partial overlap in structures is likely to lead to CLI.

Conclusion

To conclude, the aim of this study was to investigate L2 attrition in receptive and productive knowledge of VN collocations among Arabic-English returnees who had lived in the US for an extended period of time and returned at different ages to their homeland, SA. These were compared to HSs who were living in the US at the time of data collection and had not returned to SA. Our study is among the first to investigate L1 impact on processing of L2 VN collocations, measuring accuracy and reaction times. This study contributed additional evidence with respect to the need for collocational knowledge to stabilize in HSs. The study showed evidence for attrition in receptive skills among returnees in accuracy. It suggests that the L1 may have been more activated during the receptive task, resulting in Arabic influence. The productive task, however, did not show any evidence of crosslinguistic influence from Arabic, perhaps explaining why no difficulties were observed. It was also found that returnees who lost L2 input in their early childhood years were affected by CLI, whereas returnees who returned during their adolescent years are influenced by what is referred to as CLO. The findings revealed it is important to highlight that the amount of contact with both languages impacted the degree of attrition. The results also indicate that general language dominance measured with the BLP as well as vocabulary knowledge dominance differed considerably by group. In this study, due to lack of space, we cannot discuss which background variables may have affected outcomes (see Alraddadi & Treffers-Daller, in prep.).

According to the findings, attrition is adaptable to changes in input and affects the *processing* of MWUs rather than the *representations*, because participants were generally able to produce the collocations in the productive task which would not have been possible without corresponding representations. In other words, attrition does not necessarily erase or alter the underlying mental representations, but it can affect how they are processed or used. However, such a distinction between representation and processing remains problematic. It should be noted that the term attrition, in most cases, refers to online processing rather than a sign of structural deterioration (Schmid & Köpke, 2017). Thus, attrition affects the cognitive performance rather than the language knowledge itself (Paradis, 2007). Bilinguals usually encounter difficulties in lexical access even after a relatively a short period of exposure or immersion (Schmid & Köpke, 2017). The findings therefore suggest that child returnees who had been re-immersed in their L1 setting experience more processing difficulties and slower access due to the lack of L2 exposure and the need to strongly inhibit the non-target language (L1) when using L2. The study has also found that AoR is more important than LoR in the home country, since the RT1 had not acquired the MWUs to the same extent as the RT2, and their language knowledge had not stabilized sufficiently. Conversely, the RT2 had spent more of their adolescence in the US, so their knowledge had stabilized sufficiently, making them somewhat less vulnerable to language attrition.

Thus, this study provides some evidence that not every bilingual is an attriter, because there is little evidence for attrition among the RT2, whose performance on tasks was similar to the AHS. A limitation of the current study could be that attrition of L2 English may be less prominent compared to other languages, such as German, as shown in Flores's (2010) study. English, as a global language with widespread international use, offers constant exposure and opportunities for language maintenance, whereas the complexity of German morpho-syntax may present additional challenges, potentially contributing to a greater susceptibility to attrition. Nevertheless, the main concern lies in how likely is it that a returnee attrites in English upon return. The fact that English is widely spoken throughout the world is unlikely to be the key explanatory variable, since people in SA have the option to speak English if they wish, but English is not widely used. Therefore, it depends on the extent to which individuals use English in everyday life. Future research should examine these individual differences in more detail. An in-depth analysis of the complex interaction of extralinguistic factors, such as AoR, LoR, language attitudes and language use on returnees' language development, is needed to further our understanding of this particular bilingual population (see Alraddadi & Treffers-Daller, in prep.).

Abbreviations

AHS	Adolescent Heritage Speakers
AoR	Age of Return
AJT	Acceptability Judgement Task
ATH	Activation Threshold Hypothesis
BLP	Bilingual Language Profile
BNC	British National Corpus
CLI	Crosslinguistic Influence
CLO	Crosslinguistic Overcorrection
COCA	Corpus of Contemporary American English
EFL	Learning English as a Foreign Language
GFT	Gap-filling task
HL	Heritage Language
HS	Heritage Speaker
HS1	Child Heritage Speakers
HS2	Adolescent Heritage Speakers
L1	Community/first language
L2	Second Language
LoR	Length of Residence
MWUs	Multiword Units
NSs	Native speakers
PPVT	Peabody Picture Vocabulary Task
RT1	Child Returnees
RT2	Adolescent Returnees
SA	Saudi Arabia
US	United States
VN	Verb-noun

Supplementary material. To view supplementary material for this article, please visit http://doi.org/10.1017/S1366728924000610.

References

Alharthi, T. (2015). Adding More Fuel to the Fire: A study of attrition in formulaic sequences by adult learners. *Arab World English Journal (AWEJ)*, **6**(3), 230–243.

Alraddadi, H., & **Treffers-Daller, J.** (in prep.). Returnees are on the native speaker continuum too! A study of individual differences in L2 attrition.

Anderssen, M., **Lundquist, B.**, & **Westergaard, M.** (2018). Cross-linguistic similarities and differences in bilingual acquisition and attrition: Possessives and double definiteness in Norwegian heritage language. *Bilingualism: Language and Cognition*, **21**(4), 748–764. https://doi.org/10.1017/s1366728918000330

Arnon, I., **McCauley, S. M.**, & **Christiansen, M. H.** (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword

phrases. *Journal of Memory and Language*, **92**, 265–280. https://doi.org/10.1016/j.jml.2016.07.004

Barr, D. J., **Levy, R.**, **Scheepers, C.**, & **Tily, H. J.** (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, **68**(3), 255–278.

Bates, D., **Mächler, M.**, **Bolker, B.**, & **Walker, S.** (2015). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Bialystok, E., **Craik, F. I.**, & **Luk, G.** (2012). Bilingualism: consequences for mind and brain. *Trends in cognitive sciences*, **16**(4), 240–250.

Birdsong, D., **Gertken, L. M.**, & **Amengual, M.** (2012). *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*.(COERLL, University of Texas at Austin. Web. 20 Jan. 2012). https://sites.la.utexas.edu/bilingual/

Boers, F., **Demecheleer, M.**, **Coxhead, A.**, & **Webb, S.** (2014). Gauging the effects of exercises on verb–noun collocations. *Language Teaching Research*, **18**(1), 54–74.

Brehmer, B., & **Treffers-Daller, J.** (2020). *Lost in transmission: The Role of Attrition and input in heritage language development* (Vol. **59**). John Benjamins Publishing Company.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 711–733.

Bylund, E. (2019). *Age effects in language attrition.* The Oxford handbook of Language Attrition(Oxford University Press).

Davies, M. (2008). The Corpus of Contemporary American English (COCA): 560 million words, 1990-present. In *Available online at https://corpus.byu.edu/coca/*.

De Houwer, A. (2023). The danger of bilingual–monolingual comparisons in applied psycholinguistic research. *Applied Psycholinguistics*, **44**(3), 343–357. https://doi.org/10.1017/s014271642200042x

de Leeuw, E., **Tusha, A.**, **Zhao, H.**, **Helke, K.**, & **Greenfield, A.** (2018). A case of extreme phonetic attrition in the German rhotic. *Mind Matters in SLA. Edited by* Clare Wright, Thorsten Piske and Martha Young-Scholten. Bristol: Multilingual Matters, 162–183.

Dörnyei, Z., **Durow, V.**, & **Zahran, K.** (2004). Individual differences and their effects on formulaic sequence acquisition. *Formulaic sequences*, 87–106.

Dunn, L. M., & **Dunn, L. M.** (2007). *Peabody picture vocabulary test*. ((PPVT-4) (4th ed.).). (Minneapolis: Pearson.)

Flores, C. (2010). The effect of age on language attrition: Evidence from bilingual returnees. *Bilingualism: Language and Cognition*, **13**(4), 533–546.

Flores, C. (2019). Attrition and Reactivation of a Childhood Language: The Case of Returnee Heritage Speakers. *Language Learning*, **70**(S1), 85–121. https://doi.org/10.1111/lang.12350

Flores, C., **Zhou, C.**, & **Eira, C.** (2022). "I no longer count in German." On dominance shift in returnee heritage speakers. *Applied Psycholinguistics*, **43**(5), 1019–1043. https://doi.org/10.1017/s0142716422000261

González-Fernández, B., & **Schmitt, N.** (2020). Word Knowledge: Exploring the Relationships and Order of Acquisition of Vocabulary Knowledge Components. *Applied Linguistics*, **41**(4), 481–505. https://doi.org/10.1093/applin/amy057

Graham, C. R. (2012). Vocabulary attrition in adult speakers of Spanish as a second language. *Second language acquisition abroad: The LDS missionary experience*, **45**, 135.

Granena, G., & **Long, M. H.** (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second language research*, **29**(3), 311–343.

Gürel, A. (2004). Selectivity in L2-induced L1 attrition: a psycholinguistic account. *Journal of Neurolinguistics*, **17**(1), 53–78.

Gyllstad, H., & **Wolter, B.** (2016). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning*, **66**(2), 296–323.

Hunston, S. (2022). *Corpora in applied linguistics*. Cambridge University Press.

Jou, Y.-J., **Huang, C.-C. L.**, & **Cho, H.-J.** (2014). A VIF-based optimization model to alleviate collinearity problems in multiple linear regression. *Computational Statistics*, **29**, 1515–1541.

Kopotev, M., **Kisselev, O.**, & **Polinsky, M.** (2020). Collocations and near-native competence: Lexical strategies of heritage speakers of Russian. *International Journal of Bilingualism*. https://doi.org/10.1177/1367006920921594

Kubota, M., **Chevalier, N.**, & **Sorace, A.** (2020a). Losing access to the second language and its effect on executive function development in childhood: The case of 'returnees'. *Journal of Neurolinguistics*, **55**. https://doi.org/10.1016/j.jneuroling.2020.100906

Kubota, M., **Heycock, C.**, **Sorace, A.**, & **Rothman, J.** (2020b). Cross-Linguistic Influence on L2 Before and After Extreme Reduction in Input: The Case of Japanese Returnee Children. *Front Psychol*, **11**, 560874. https://doi.org/10.3389/fpsyg.2020.560874

Kupisch, T. (2014). Adjective placement in simultaneous bilinguals (German–Italian) and the concept of cross-linguistic overcorrection. *Bilingualism: Language and Cognition*, **17**(1), 222–233. https://doi.org/10.1017/s1366728913000382

Kupisch, T., **Kolb, N.**, **Rodina, Y.**, & **Urek, O.** (2021). Foreign Accent in Pre- and Primary School Heritage Bilinguals. *Languages*, **6**(2). https://doi.org/10.3390/languages6020096

Larson-Hall, J. (2019). L2 Lexical Attrition. *The Oxford handbook of Language Attrition*.

Laufer, B. (1989). 25 What Percentage of Text-Lexis is Essential for Comprehension? *Special language: From humans thinking to thinking machines*, 316.

Laufer, B., & **Waldman, T.** (2011). Verb-noun collocations in second language writing: A corpus analysis of learners. *' English.* Language Learning, **61**(2), 647–672.

Martinez, R., & **Schmitt, N.** (2012). A phrasal expressions list. *Applied Linguistics*, **33**(3), 299–320.

Masrai, A., & **Milton, J.** (2019). How many words do you need to speak Arabic? An Arabic vocabulary size test. *The Language Learning Journal*, **47**(5), 519–536.

Matos, J., & **Flores, C.** (2022). More insights into the interaction between age, exposure, and attitudes in language attrition and retention from the perspective of bilingual returnees. *International Journal of Bilingualism*, 13670069221136941.

Mehotcheva, T. H., & **Köpke, B.** (2019). Introduction to L2 attrition. *The Oxford handbook of Language Attrition*.

Mehotcheva, T. H., & **Mytara, K.** (2019). Exploring the impact of extralinguistic factors on L2/FL attrition.

Montrul, S. (2016). *The acquisition of heritage languages*. Cambridge University Press.

Müller, N., & **Hulk, A.** (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition*, **4**(1), 1–21.

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied Linguistics*, **24**(2), 223–242.

Nguyen, T. M. H., & **Webb, S.** (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, **21**(3), 298–320.

Nicklin, C., & **Plonsky, L.** (2020). Outliers in L2 research in applied linguistics: A synthesis and data re-analysis. *Annual Review of Applied Linguistics*, **40**, 26–55.

Paradis, M. (2007). L1 attrition features predicted by a neurolinguistic theory of bilingualism. *Language attrition: Theoretical perspectives*, **33**, 121–133.

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, **162**(1–2), 8–13.

Pulido, M. F. (2022). Why are multiword units hard to acquire for late L2 learners? Insights from cognitive science on adult learning, processing, and retrieval. *Linguistics Vanguard*, **8**(1), 237–247. https://doi.org/10.1515/lingvan-2021-0043

Rothman, J., & **Treffers-Daller, J.** (2014). A Prolegomenon to the Construct of the Native Speaker: Heritage Speaker Bilinguals are Natives Too! *Applied Linguistics*, **35**(1), 93–98. https://doi.org/10.1093/applin/amt049

Sá-Leite, A. R., **Flores, C.**, **Eira, C.**, **Haro, J.**, & **Comesaña, M.** (2023). Language balance rather than age of acquisition: A study on the cross-linguistic gender congruency effect in Portuguese–German bilinguals. *Bilingualism: Language and Cognition*, 1–14. https://doi.org/10.1017/s1366728923000378

Schmid, M. S., & **Jarvis, S.** (2014). Lexical access and lexical diversity in first language attrition. *Bilingualism: Language and Cognition*, **17**(4), 729–748. https://doi.org/10.1017/s1366728913000771

Schmid, M. S., & **Köpke, B.** (2017). When is a bilingual an attriter? Response to the commentaries. *Linguistic Approaches to Bilingualism*, **7**(6), 763–770.

Sonbul, S., & **El-Dakhs, D.** (2020). Timed versus untimed recognition of L2 collocations: Does estimated proficiency modulate congruency effects? *Applied Psycholinguistics*, **41**(5), 1197–1222. https://doi.org/10.1017/s0142716420000051x

Szudarski, P. (2012). Effects of meaning-and form-focused instruction on the acquisition of verb-noun collocations in L2 English. *Journal of Second Language Teaching & Research*, **1**(2), 3–37.

Team, R. C. (2013). R: A language and environment for statistical computing [R]. R Foundation for Statistical. Computing http://www.R-project.org/.

Tomiyama, M. (2009). Age and proficiency in L2 attrition: Data from two siblings. *Applied Linguistics*, **30**(2), 253–275.

Treffers-Daller, J., **Daller, M.**, **Furman, R.**, & **Rothman, J.** (2016). Ultimate attainment in the use of collocations among heritage speakers of Turkish in Germany and Turkish–German returnees. *Bilingualism: Language and Cognition*, **19**(3), 504–519.

Wechsler, D. (2008). *Wechsler adult intelligence scale–Fourth Edition (WAIS–IV)*. San Antonio, TX: NCS Pearson, **22**(498), 1.

Wolter, B., & **Yamashita, J.** (2015). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, **36**(5), 1193–1221.

Wolter, B., & **Yamashita, J.** (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance? *Studies in Second Language Acquisition*, **40**(2), 395–416.

Wood, D. (2019). Classifying and Identifying Formulaic Language. In *The Routledge handbook of vocabulary studies* (pp. 30–45). https://doi.org/10.4324/9780429291586-3

Yamashita, J., & **Jiang, N.** (2010). L1 Influence on the Acquisition of L2 Collocations: Japanese ESL Users and EFL Learners Acquiring English Collocations. *TESOL Quarterly*, **44**(4), 647–668. https://doi.org/10.5054/tq.2010.235998

Zaabalawi, R. S. (2019). English collocations versus Arabic collocations: A pedagogical dimension. *The Reading matrix: An international online journal*, **19**(1), 167–180.

https://doi.org/10.1017/S1366728924000610 Published online by Cambridge University Press