





Individual differences in foreign language attrition: a 6-month longitudinal investigation after a study abroad

Anne Mickan ^{a,b}, James M. McQueen ^{a,c}, Laurel Brehm ^c and Kristin Lemhöfer ^a

^aDonders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands; ^bInternational Max Planck Research School for Language Sciences, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands; ^cMax Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

ABSTRACT

While recent laboratory studies suggest that the use of competing languages is a driving force in foreign language (FL) attrition (i.e. forgetting), research on “real” attriters has failed to demonstrate such a relationship. We addressed this issue in a large-scale longitudinal study, following German students throughout a study abroad in Spain and their first six months back in Germany. Monthly, percentage-based frequency of use measures enabled a fine-grained description of language use. L3 Spanish forgetting rates were indeed predicted by the quantity and quality of Spanish use, and correlated negatively with L1 German and positively with L2 English letter fluency. Attrition rates were furthermore influenced by prior Spanish proficiency, but not by motivation to maintain Spanish or non-verbal long-term memory capacity. Overall, this study highlights the importance of language use for FL retention and sheds light on the complex interplay between language use and other determinants of attrition.

ARTICLE HISTORY



Received 26 August 2020
Accepted 29 April 2022


KEYWORDS

Foreign language attrition;
language use; study abroad;
individual differences;
longitudinal study

In the globalised world we live in, more and more people enrol in language classes or move abroad for extended periods of time: the EU-funded Erasmus programme alone has sent out more than 10 million people to other countries since its establishment in 1987 (European Commission, 2020). Along with the cultural enrichment and personal growth that comes with experience abroad, most people hope to improve their foreign language (FL) skills. Maintenance of once acquired language skills, however, often proves difficult once the new language is no longer used regularly. According to anecdotal reports, after the end of a language course or stay abroad, learners soon start losing their FL skills, especially vocabulary (but see, for instance, Hessel, 2020; Huensch et al., 2019, for reports of retained aspects of proficiency after return). A healthy speaker’s gradual loss of a language across time due to forgetting is also referred to as *language attrition* (e.g. Schmid & Mehotcheva, 2012), with vocabulary loss as one of its most instantaneous and notable aspects (e.g. Cohen, 1989). How come we forget foreign languages and what determines how fast and how much we do so?

Recent experimental research has shown that forgetting of FL vocabulary can come about through the use (or the learning) of other languages, especially other foreign languages (e.g. Bailey & Newman, 2018; Isurin & McDonald, 2001; Levy et al., 2007; Mickan et al., 2020). Mickan et al. (2020), for example, showed that naming pictures in L2 English or L1 Dutch hampers subsequent access to recently learned L3 Spanish translation equivalents. Merely by having participants retrieve translations in another language, Mickan and colleagues thus succeeded in inducing language access problems in the lab and in doing so, established language interference as one possible driving force in FL forgetting. If these lab-based results generalise to real life contexts, language usage patterns during the attrition period should be crucial determinants of the rate and extent of target foreign language loss in natural attriters. Paradoxically though, this prediction has not been borne out in previous research of foreign language attrition “in the wild” (Bahrick, 1984a, 1984b; Hessel, 2020; Mehotcheva & Mytara, 2019; Xu, 2010). Have previous studies with real attriters failed to accurately assess and document the role of language

CONTACT Anne Mickan  a.mickan@donders.ru.nl  Donders Institute for Brain, Cognition and Behaviour, PO Box 9104, Nijmegen 6500 HE, The Netherlands

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/23273798.2022.2074479>.

This article has been republished with minor changes. These changes do not impact the academic content of the article.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

use in foreign language attrition, or does language use simply play a much smaller role in real life than in tightly controlled lab situations? The present study seeks an answer to this question by looking at language forgetting following highly immersive L2 learning during a study abroad (for a discussion of research into the effects of studying abroad on L2 proficiency, see, e.g. Tullock & Ortega, 2017; Yang, 2016). Supported by regular frequency of use ratings, we asked whether there indeed is a relationship between language use and foreign language forgetting in the first six months after a study abroad, and whether there is an accessibility trade-off between the languages that an attriter speaks.

Longitudinal vs cross-sectional designs

Previous research on foreign language attrition in real life can be divided into longitudinal and cross-sectional studies. The longitudinal approach is the most intuitive way of studying the phenomenon: here, researchers follow a group of attriters over a period of time, assessing their FL skills at regular intervals. This approach, however, is very time-consuming. Furthermore, testing participants at multiple timepoints constitutes practice and may therefore prevent attrition, counteracting the very phenomenon researchers are after. Participants also tend to drop out along the way, resulting in often small, under-representative participant samples (e.g. $N = 5$ in Mehotcheva, 2010; $N = 2$ in Tomiyama, 2008; $N = 4$ in Yoshitomi, 1999; though see Murtagh, 2003; and Xu, 2010, for exceptions). To circumvent these issues with data collection, many studies on FL attrition have used cross-sectional designs instead (e.g. Abbasian & Khajavi, 2010; Bahrack, 1984a, 1984b; Hansen & Chen, 2001), or a combination of cross-sectional and longitudinal designs (e.g. Grendel, 1993; Mehotcheva, 2010; Weltens, 1988). In cross-sectional studies, groups with differing attrition lengths are compared to each other as well as to a baseline group of non-attriters (i.e. a group of comparable learners) of the foreign language. Although cross-sectional studies have provided valuable insights, statistical comparisons in such studies rely on the groups being matched on a multitude of factors, such as age, socio-economic status, FL learning context, and the level of FL proficiency reached prior to attrition onset. While matching on the first three might be feasible, variation in FL proficiency is difficult to control. Attriters who have reached different levels of proficiency in the foreign language will attrite at different rates (e.g. Bahrack, 1984a, 1984b), making it important to account for initial proficiency in interpreting later attrition. In cross-sectional studies, however, it

is extremely difficult to accurately estimate prior FL proficiency on an individual basis. For a thorough investigation into individual differences in foreign language attrition, we thus instead need large-scale longitudinal studies with a pre-attrition baseline measure next to the attrition measurement itself. A comparison between performance on the dependent measure (e.g. a vocabulary or grammar test) at baseline and at the attrition measurement (i.e. after a period of disuse) can then provide accurate and fine-grained participant-specific forgetting rates, which can serve as the basis for individual difference analyses. The present study aims at providing such a large-scale longitudinal dataset.

Language use as a predictor of foreign language attrition

From a theoretical point of view, researchers in the domain of (foreign) language attrition unanimously agree that language use should be one of the key factors in language attrition: for vocabulary maintenance at least, continued use is understood to be necessary to keep the activation thresholds of words in a given language low, and use of languages other than the target language (i.e. competing lexical entries) is thought to increase these activation thresholds and hence complicate subsequent retrieval (the Activation Threshold Hypothesis, Köpke, 2002; Paradis, 2004). Both theory and lab-based experimental evidence thus point towards a clear role for language use and interference in attrition, and yet, as Mehotcheva and Köpke (2019) summarise, the majority of studies that have investigated the role of language use report no (consistent) relationship between language use (of the target and / or other languages) and foreign language retention (Bahrack, 1984a, 1984b; Hessel, 2020; Mehotcheva, 2010; Xu, 2010; but see also Huensch et al., 2019).

Obviously, in real life, using a foreign language is inversely related to using other languages: code-switching aside, you only ever use one language at a time, and the more time you spend using a given language, the less time you have left for using other languages. Hence, what studies care about the most is usually only target language use (though use of other languages is often asked for as well, see for example Mehotcheva, 2010). In a large-scale cross-sectional study on the retention of school-/university-learned L2 Spanish, for example, Bahrack (1984a, 1984b) found none of four Spanish language use measures (reading, writing, listening and speaking) to correlate with Spanish retention. Likewise, Mehotcheva (2010) failed to find clear evidence for a relationship between Spanish language use and Spanish vocabulary retention

in German and Dutch learners of Spanish in the initial months after a study period abroad. Likewise, German/Dutch and English use did not predict Spanish attrition either. An exception to this pattern of results is a study by Alharthi and Al Fraidan (2016), who found that L2 English internet usage by L1 Arabic participants was a good predictor of L2 English proficiency after a 15-month attrition period. Yet, even in the latter study, other target language frequency of use measures, which should intuitively be just as important (such as watching TV in English, reading books, or attending English FL courses), did not predict L2 proficiency either.

What causes the failure to document a consistent, beneficial role of target language use, or conversely, a detrimental role of non-target language use on retention? Bahrck (1984a, 1984b) reasoned that there was not enough variance in terms of Spanish use in his sample: all of his participants used very little to no Spanish during the attrition period. Others have argued that frequency of use measures often fail to accurately describe and capture language use because they focus too much on quantity (e.g. absolute hours of use) as compared to the quality of the input (see Fernández & Gates Tapia, 2016; Schmid, 2007). Another possible problem might be how language use is quantified. Especially in cross-sectional studies, such as the ones by Bahrck (1984a, 1984b) and Mehotcheva (2010), participants are asked to judge their frequency of use in *retrospect*. Estimating the overall amount of exposure to a language for a long period of time (more than 12 months for some of the participants in Mehotcheva's study and up to 50 years in Bahrck's sample) is difficult and prone to over- or underestimation. Moreover, judgments are often given on relatively subjective scales: Mehotcheva (2010), for example, had participants judge frequency on a 5-point-scale from "very rarely" to "very frequently". Taking these two aspects and the above discussed baseline problem for cross-sectional studies together, it is thus possible that frequency of use measures in previous studies were simply too imprecise to predict changes (or group differences) in proficiency. In the present study, we instead collected multiple, percentage-based frequency of use measures at regular intervals to circumvent this issue (see next section for details).

The current study

To assess the role of language use in foreign language attrition, we followed a large group ($N=97$) of L1 German learners of L3 Spanish over the course of a year, spanning both their one-semester-long study abroad in Spain as well as their first half a year back in Germany. We chose to assess attrition rates after six

months, because previous research by Bahrck (1984a, 1984b) suggests that most forgetting happens within the first months and years after exposure to the foreign language stops. For feasibility reasons and to reach a large number of potential participants, we tested participants online. To assess changes in Spanish proficiency over time, we administered an online picture-naming vocabulary test in Spanish just before the study abroad (T1), at the end of the study abroad (T2) and roughly six months post return to Germany (T3). Since this study is about forgetting, the analyses reported below concern changes from T2 to T3. The T1 measurement is only relevant for two secondary analyses, as explained further below. We chose to study productive vocabulary for maximal comparability with the above-cited lab-based language attrition studies, and because productive FL skills are known to attrite first (e.g. Bahrck, 1984a, 1984b), increasing our chances of observing attrition in the six-month attrition window. In the remainder of this paper, terms such as Spanish proficiency and Spanish attrition always refer to Spanish *lexical* proficiency and Spanish *lexical* attrition, respectively, even when we do not explicitly specify this every time.

Next to the Spanish proficiency test, we administered fluency tests in German (L1) and English (L2) and a language background questionnaire at each session, as well as a short frequency of use questionnaire once every month in between. To get representative and accurate frequency of use estimates, we opted for a more continuous and less abstract frequency of use measure than previous studies. Instead of asking our participants once in retrospect, we asked them to estimate their *current* frequency of use *once every month* during the attrition period. This resulted in about six frequency of use measurements to average over rather than one single measurement. As in Mehotcheva (2010), we asked for frequency of use indications in the target foreign language Spanish, as well as in L1 German, L2 English and any other languages. Rather than making judgments on a scale, we asked our participants to estimate the percentage of time they currently spent speaking (and listening, reading and writing) each language. The total for each of these four domains had to add up to 100%, such that a given percentage reflected, for example, the percentage of time someone currently spoke Spanish out of the total amount of time they spent speaking. Because these percentages were given relative to a participant's personal total amount of language use (i.e. their 100%) rather than some subjectively perceived notion of "rare" and "frequent", we expected that they would provide a more reliable estimate than indications on a scale or estimates given in hours or minutes.

As already noted, amount of target FL use and the use of other languages are inversely related, and so finding a positive relationship between Spanish use and Spanish retention entails finding a negative relationship between Spanish retention and the amount of use of all other languages. An interesting question, however, is whether it matters *which* other languages an individual speaks. Laboratory results from Mickan et al. (2020) suggest that other foreign languages interfere more with a new foreign language than a native language does. Next to asking whether more frequent Spanish use during the attrition period helps Spanish retention, we thus also asked whether the ratio of L2 English over L1 German use during the remaining time makes a difference: does someone who predominantly uses German when they do not speak Spanish suffer less from attrition than someone who predominantly uses English?

In order to investigate whether there is a trade-off in accessibility between languages, we had to complement the (partly mutually dependent) frequency of use ratings. To do so, we administered fluency tests in L1 German and L2 English and asked whether changes in Spanish proficiency would go hand in hand with changes in fluency in these other two languages. If, as lab studies suggest, there is such a trade-off (caused through interference between languages) and assuming that frequent language use results in higher fluency scores, we should observe increases in fluency in German and English to be associated with proficiency decreases in Spanish.

Finally, since quantity of input in a certain language might not be the sole, or even most important predictor of Spanish retention, we also asked for the type of input our participants received. We therefore asked for frequency of use ratings separately in reading, writing, listening and speaking. We also asked our participants to report what percent they received native as compared to non-native input in Spanish. Individuals who received solely native input, regardless of the total amount of input they get, might show less signs of attrition, or attrite more slowly than people who received mostly non-native and thus potentially incorrect input.

Other determinants of foreign language forgetting

Language use is unlikely to be the only relevant predictor of forgetting rates “in the wild”. Many factors have been implicated in foreign language attrition (for the most recent summary, see Mehotcheva & Mytara, 2019). In a naturalistic experiment, these other factors should be taken into account to arrive at a parsimonious model of foreign language attrition. In the current study, we thus additionally included a number of variables that

either have been consistently shown to impact the rate of forgetting, or that have yielded contradictory results. In accounting for those variables, we hoped to get a more complete picture of the determinants of foreign language attrition.

Motivation

Similar to (and linked with) language use, one might expect motivation to learn a foreign language well and one’s attitude towards the FL to be important in determining how well FL skills are maintained. Once again though, the empirical evidence for a role of motivation in language maintenance (and conversely loss) is sparse (see Mehotcheva & Mytara, 2019, for a recent discussion). Just as with language use, previous studies assessed motivation only once, at the attrition measurement (i.e. once attrition had already occurred). Motivation before attrition onset, however, is arguably at least as important in determining the effort someone will put into maintaining a foreign language. Moreover, motivation is dynamic and can change over time (Nikitina & Furuoka, 2005), and hence can differ before and after attrition onset. For a more nuanced and complete picture, we administered a shortened version of Gardner’s Attitude and Motivation questionnaire (see Methods) at each of the main measurement sessions. We then averaged over the pre-attrition (T2) and attrition (T3) measurements to arrive at an estimate of overall motivation and attitude towards Spanish during the attrition period and asked whether this estimate predicts forgetting rates.

Amount of experience with the foreign language

Length of exposure to the foreign language and thus amount of experience with it prior to attrition onset is also often thought to be important. Usually, length of exposure is operationalised as the length of the stay abroad. In our case, all participants went abroad for only one semester. The variance in terms of study abroad length is thus minimal in our sample and most likely not meaningful in itself. In a comparable population of German and Dutch learners of Spanish, Mehotcheva (2010) also found no conclusive evidence for a role of study abroad length on Spanish language retention, even though in her sample the range of the study abroad period was almost twice that in our sample. In adult FL learners who only go abroad for a short period of time, the study abroad is also often only the tip of the iceberg. For many exchange students, much of the learning of the FL happens before they go abroad. It is hence variation in the amount (and possibly intensity) of experience with Spanish prior to the study abroad that we think is most important for a sample

like the one tested in our study. Our learners started their study abroad with different amounts of Spanish experience and hence we asked whether people with more years of experience were less prone to undergo attrition after the study abroad than those who had started learning Spanish more recently. The intensity of their prior experience with Spanish might of course also play a role, but we are not aware of a standardised way of quantifying exposure intensity nor of calculating a composite score with length of exposure and hence decided to only take years of experience into account.

Foreign language proficiency before attrition onset

A variable closely linked to the amount of FL experience is the proficiency level reached in the FL prior to attrition onset. In fact, even though it has been claimed to be less important than the amount of FL experience (Hansen, 1999), it is the variable that has been linked most consistently to forgetting rates: People with higher levels of FL attainment before attrition onset have repeatedly been shown to suffer relatively less from attrition than participants with lower levels of pre-attrition proficiency (e.g. Bahrick, 1984a, 1984b; Mehotcheva, 2010; Murtagh, 2003; Weltens, 1988). Bahrick (1984a, 1984b), for example, found that participants who had followed more FL courses and who had received higher course grades prior to attrition onset maintained a higher proportion of vocabulary in what he called “permastore” (i.e. vocabulary that remains available to a language user even after more than 25 years of disuse of the FL). While intuitively FL attainment should correlate with the amount of FL experience, this need not necessarily be the case: some participants might be more efficient and faster learners than others and hence achieve higher levels of proficiency in a shorter amount of time. It thus seemed important to include both FL experience and FL proficiency in our analysis.

Long-term memory capacity

A factor that we know very little about in relation to (foreign) language attrition is long-term memory capacity. Most recent theoretical accounts of (FL) attrition draw links between language and domain-general forgetting, and highlight the possibility of common underlying neural substrates and cognitive processes (see for example, Ecke, 2004; Köpke & Keijzer, 2019; Linck & Kroll, 2019; Mickan et al., 2019). If the same or similar mechanisms underlie both domain-general (i.e. non-verbal) and language forgetting, it may be that someone’s rate of FL attrition is partially determined by their (non-verbal) long-term memory capacity. We tested this by administering a standardised visual long-term memory test at T3 (the Doors test; Baddeley et al., 1994, 2016). To the best of our

knowledge, we are the first to investigate whether non-verbal long-term memory capacity predicts foreign language attrition severity.

Attrition self-judgment

Finally, we asked whether participants have a realistic perception of how much they forget. Previous research suggests that participants tend to overestimate their personal amount of attrition (e.g. Murtagh, 2003; Weltens, 1988). While this is not a variable that is thought to predict or cause attrition, it is nevertheless interesting to ask whether we observe a similar overestimation of foreign language loss in our population.

Item-specific factors: word frequency and cognate status

Next to individual difference factors, there are also variables that may influence the rate of forgetting on the item level. High frequency words, for example, have been found to be retained better than low frequency words (e.g. Mehotcheva, 2010), and cognates tend to be remembered better than non-cognates (e.g. Weltens, 1988; though see Engstler, 2012). We therefore included both of these factors in the analysis.

Learning context, age and other constants in the present study

Of course, the above discussed variables do not constitute an exhaustive list of the factors contributing to foreign language forgetting. Including all possible predictors in one model was beyond the scope of the current study, and so, in an effort to reduce the number of predictors, we kept some factors constant. These factors include the learning context (natural / immersion or instructed), the languages involved, as well as the age and socio-economic status of the attriters and the length of the attrition period (see Mehotcheva & Köpke, 2019; Mehotcheva & Mytara, 2019, for overviews and discussions of these variables). We selected our participants to be as closely matched on these aspects as possible: all of our participants were German university students between 20 and 29 years of age. They all went to Spain on a study abroad and hence learned Spanish both under natural circumstances while abroad as well as in a formal (classroom) setting prior to the study abroad. We scheduled the attrition session to take place roughly six months after the end of the study abroad for everyone, and participants were all back in Germany and immersed in their L1 at that time point. Six months is a relatively short attrition period. Previous research, however, suggests that most forgetting happens within the first months and years

after attrition onset (Bahrick, 1984a, 1984b). We thus expected to observe at least some amount of attrition within six months. Investigating a much longer period than that was simply not possible given the time constraints of the PhD project that this research was part of.

Attrition as the mirror image of acquisition?

The fact that we have a pre-study-abroad baseline (T1) next to the pre-attrition baseline (T2) gave us the opportunity to ask an additional question about the nature of attrition: namely whether the forgetting process is the reverse of the acquisition process, and hence whether the information acquired last is also the first to be forgotten. This hypothesis was originally formulated by Jakobson (1941) to explain pathological L1 loss and is commonly known as the Regression Hypothesis (RH). Since its initial formulation, alternative versions have been advocated and researchers have argued that it might not be the information learned last, but rather the information learned *least well* that is forgotten first (e.g. Hedgcock, 1991). The jury on this debate is still out, partially because these two versions of the RH are hard to tease apart in real life, where information learned last often is also learned least well. In fact, there is evidence in support of the RH in both of its formulations (e.g. Hansen, 1999; Hedgcock, 1991; Kuhberg, 1992; Olshtain, 1989). Most of this evidence, however, comes from studies with children and from the domain of syntax and morphology. For vocabulary and late adult L2 learners, to the best of our knowledge, the RH has only been tested (and confirmed) once, in a cross-sectional design (Wang, 2010). There is thus still a need for more research on whether the words learned last (or least well) are indeed the first to be forgotten. Having not only a pre- and post-attrition (i.e. T2 and T3) measure, but also a pre-study abroad baseline (T1) enabled us to ask, in an additional analysis, whether the words learned most recently (i.e. during the study abroad: unknown at T1 but known at T2), are more likely to be forgotten by T3 than the words that were already known before the study abroad (i.e. known at both T1 and T2). This does not distinguish the two versions of the RH from one another, but it does test the RH in its original formulation.

Methods

Participants

During the summer months of 2018, we invited students of 33 German universities who were about to embark on a study period in Spain to our study. In total, we received 481 sign-ups, of which we selected 194 German native speakers. Selection criteria were having German as

their only mother tongue, the length of the study abroad period (4–7 months), their study abroad start date (no earlier than mid-August 2018, no later than October 2018) and whether they were planning on attending university courses in Spanish or English (see Supplementary Materials, S1, for details).

Due to participant drop-out and technical difficulties with the online recording of audio responses in some of the tasks, only 99 of these participants contributed data to both T2 and T3 (see Supplementary Materials, S1, for details on drop-out rates). Next to exclusions based on data availability, we excluded an additional two participants because of possible distraction or cheating during the Spanish oral naming task (indicated by frequent typing noises). The remaining 97 participants form the dataset for all analyses reported in this paper.

Participants from the final set (71 females) were between 20 and 29 years of age ($M = 22.30$, $SD = 1.80$),

Table 1. Participant characteristics.

	<i>M</i>	<i>SD</i>	Range
			<i>Spanish</i>
Spanish AoA	15.82	3.24	10–26
Amount of experience with Spanish (in years, at T1)	3.78	2.32	0–11
Spanish proficiency self-ratings			T1
Speaking	3.70	1.58	1–7
Writing	3.65	1.38	1–6
Listening	4.26	1.44	1–7
Reading	4.39	1.25	1–7
Mean across domains	4.00	1.41	
			T2
Speaking	4.49	1.15	1–7
Writing	4.34	1.18	1–7
Listening	4.91	1.20	1–7
Reading	4.99	1.04	1–7
Mean across domains	4.68	1.14	
			T3
Speaking	4.30	1.29	1–7
Writing	4.09	1.15	1–6
Listening	4.78	1.20	1–7
Reading	5.08	1.03	1–7
Mean across domains	4.56	1.17	
			<i>English</i>
English AoA	8.86	1.93	3–14
English proficiency self-ratings			T1
Speaking	5.29	0.91	3–7
Writing	4.93	1.03	3–7
Listening	5.68	0.84	4–7
Reading	5.63	0.98	2–7
Mean across domains	5.38	0.94	
			T2
Speaking	5.26	0.90	3–7
Writing	4.98	1.10	2–7
Listening	5.72	0.80	3–7
Reading	5.73	0.88	3–7
Mean across domains	5.42	0.92	
			T3
Speaking	5.23	1.08	1–7
Writing	5.04	1.22	1–7
Listening	5.77	0.92	2–7
Reading	5.73	0.97	4–7
Mean across domains	5.44	1.05	

Notes: AoA = age of acquisition. Proficiency self-ratings were acquired on a 7-point Likert scale (1 = very poor, 7 = like a native speaker).

had normal or corrected-to-normal vision and reported no history of neurological impairment or speech related disabilities. For all participants, English was their first foreign language. At the time of recruitment, prior to their study abroad, all of them knew some Spanish, though to varying degrees. Proficiency self-ratings at each of the three main measurement time points, both for English and Spanish, can be inspected in Table 1. Other languages that participants knew included most prominently French and Latin. Spanish is referred to as L3 in this paper, even though it was in fact L4 or even L5 for some of our participants.

22 of the 97 participants studied Spanish / Latin American studies at their German home universities; the remaining participants came from other study programmes. Figure 1A illustrates where participants' home universities were situated. Study abroad destinations within Spain varied and can be inspected in Figure 1B. Participants went to Spain for on average 5.14 months ($SD=0.72$, range=3.53–7.96). At the time of the attrition measurement (T3), participants had been back in Germany for on average 6.17 months ($SD=0.45$, range=4.9–7.7; see Procedure section for details on session timings).

Participants took part on a voluntary basis and were reimbursed via bank transfer with €20 per completed time point, thus making for a total of €60 if they completed all three sessions. The study was approved by the ethics committee of the Faculty of Social Sciences, Radboud University (ESCW2016-1403–391).

Procedure & materials

Overview

We followed participants for one year, spanning their study abroad period and their first six months back in

Germany, thus documenting both their study abroad and the initial stages of the subsequent attrition process. Throughout this time, participants were tested online at three time points: at the beginning of their study abroad (T1), at the end (T2), and roughly six months after leaving Spain (T3), when they were back in Germany (see Figure 2). We will subsequently refer to those measurement time points as sessions. At each session, participants first completed a questionnaire, followed by a Spanish picture naming vocabulary test, and finally a number of English and German fluency tests. At T3, participants additionally completed a long-term memory test. In between these sessions, about once every month, participants were also asked to complete a questionnaire rating their current frequency of use in Spanish, English and German.

All tasks were administered online. For each session, participants received an email with personalised links to each task, in the order they had to be completed in, and the instruction to do so within two weeks. If necessary, we sent reminder emails (see Supplementary Materials, S2, for details). For all tasks, participants were asked to find a quiet spot in which to do the tasks alone, within one sitting and in the indicated order. Time stamps on the logfiles suggest that all participants complied with those instructions. Logfiles and audio recordings were stored on a secure Radboud university server to which only the first author of this paper had access.

Timings of sessions

The timing of the tests was participant-specific and depended on when a participant had started to study abroad, when they were planning to return to Germany, and whether they had any extended trips to other countries planned in between, such as trips back

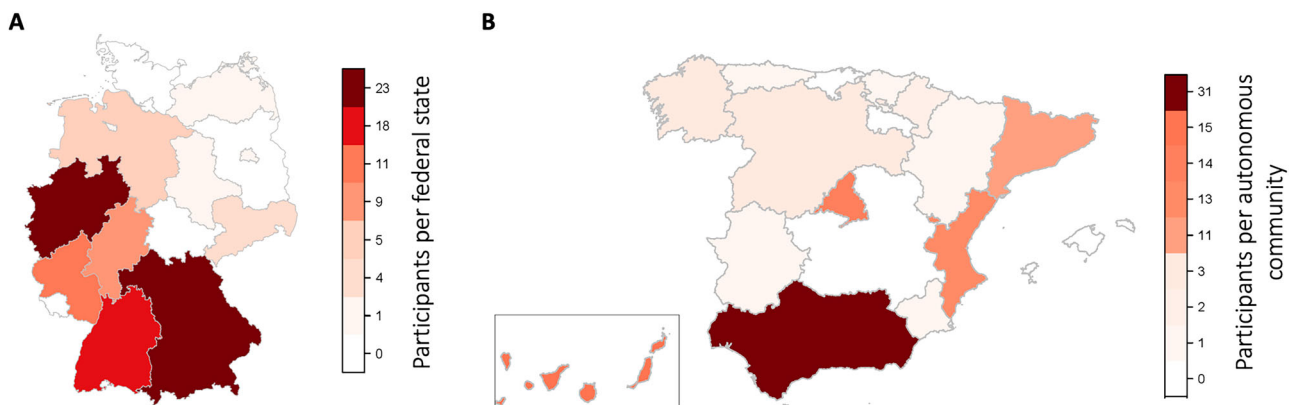


Figure 1. A. Map of Germany showing where participants' home universities were located. B. Map of Spain showing where participants went to study abroad.

to Germany between T1 and T2 and trips back to Spain between T2 and T3.

T1 timing. As soon as participants signed up and were deemed suitable for the study (see Participant section), they were invited for the first session. Participants completed T1 either before they left for Spain (10%), or within their first (45%), second (38%) or third (7%) week in Spain.

T2 timing. T2 was initially scheduled to take place two weeks before participants would return to Germany. However, because the study abroad spanned the Christmas vacation and, as determined in a short pre-Christmas questionnaire, most participants went back to Germany for the holidays, we had to reschedule T2 for some of them. Participants who went back to Germany for a week or longer and who would return to Spain for only a relatively short amount of time (see Supplementary Materials, S3, for details), were invited to complete T2 before Christmas (29 out of the 97 participants) to ensure that their T2 measurement reflected their peak Spanish performance. All remaining participants were invited as originally planned, two weeks prior to the end of their study abroad. 77% of them completed T2 while still in Spain, 19% within the first two weeks back in Germany and 4% within the first three weeks back in Germany.

T3 timing. T3 was initially scheduled exactly six months after a participant left Spain. Given, however, that these T3 dates spanned the summer vacation period, we adjusted individual T3 dates based on participants'

indicated summer vacation plans (collected in a questionnaire). Adjustments were made such that T3 would ideally take place before any vacation, but most importantly before any trips to Spanish-speaking countries, while keeping the time between T2 and T3 maximal and close to 6 months. If this was not possible, we made sure that there was at least one month in between returning from vacation and the respective T3 test. On average, T3 took place 6.18 months after participants left Spain ($SD = 0.45$, range = 4.9–7.7).

The tasks administered at each session always followed the same procedure, described as follows.

Spanish vocabulary test (dependent variable)

Materials. To make sure that our vocabulary measure was sensitive to the type of vocabulary knowledge acquired while abroad, item selection was informed by results from a pilot study with five participants who had been on a similar stay in a Spanish-speaking country. For a larger set of items, they indicated whether they knew each word and whether they thought they learned it during their stay abroad. We selected from this item set as many words as possible that tended to be learned abroad.

The resulting set of items consisted of 144 pictures of everyday objects and animals in Spanish (see the Appendix for a full list of items) that had to be named at each session. The first four of those were practice items and were not included in analyses. Three additional items were excluded because their corresponding pictures turned out to be ambiguous (e.g. the picture of a *pearl* often elicited *ball*). Out of the remaining 137

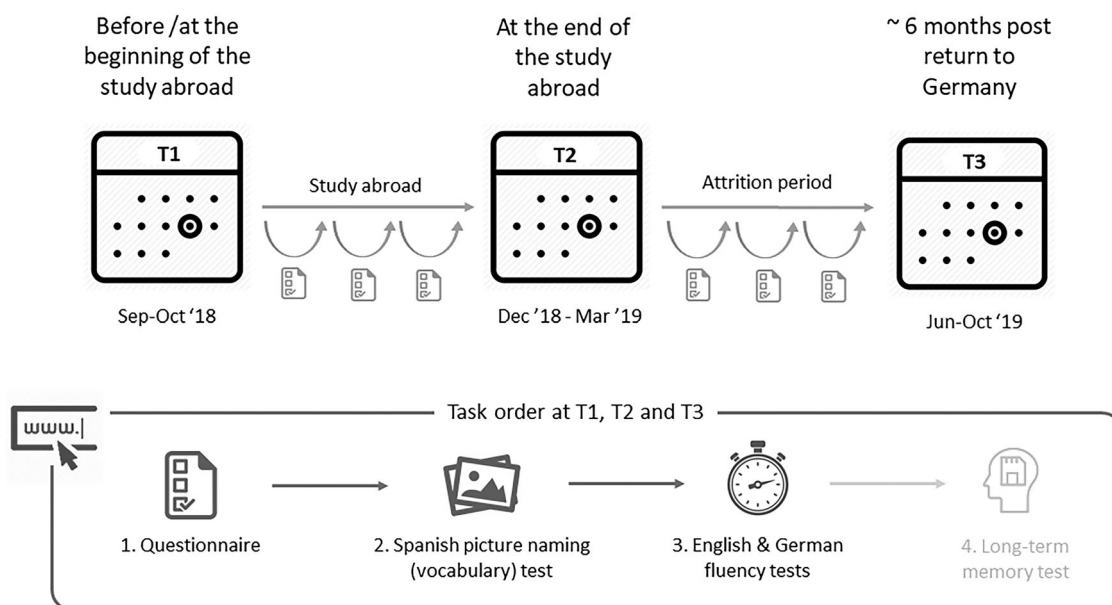


Figure 2. Online study overview.

experimental items, 18 were cognates between Spanish, German and English, 23 were cognates between Spanish and English only, and the remaining 96 words were non-cognates in the three languages. We defined cognates as translation equivalents with a form overlap of at least 50%, either phonologically or orthographically. Cognates thus included both identical cognates, such as “sofa” (German: Sofa, Spanish: sofá), and non-identical cognates, like “botella” for the English word “bottle” (see the Appendix for a full list). Because we did not see performance differences in Spanish recall between cognates in all three languages and cognates in only Spanish and English (neither at T2: $t(38.50) = 0.52$, $p = .605$, nor at T3: $t(37.88) = 0.64$, $p = .526$), we collapsed over the two types of cognates for the analyses below, distinguishing only non-cognates from cognates (of any type).

Experimental items were between 1 and 5 syllables long in Spanish ($M = 2.64$, $SD = 0.81$). Their Spanish log frequencies ranged from 1.08 to 4.67 ($M = 2.63$, $SD = 0.62$, according to the Spanish Subtlex, Cuetos et al., 2011) and their corresponding German log lemma frequencies ranged from 0.30 to 3.77 ($M = 2.29$, $SD = 0.65$, according to the German Subtlex, Brysbaert et al., 2011). We chose items from all frequency bands to make sure that participants’ performance would not reach ceiling or floor at any session. One could argue for inclusion of either the Spanish or the German frequency counts in a model on Spanish forgetting rates, which are highly correlated ($r = 0.74$). Nevertheless, for analysis, we chose the one which best predicted Spanish forgetting rates, which turned out to be Spanish log frequency (see Supplementary Materials, S8).

Pictures were photographs taken from Google images and the BOSS database (Brodeur et al., 2010). All pictures were displayed on a white background and occupied a maximum of 400 px in either width or length.

Procedure. After passing a microphone detection test, participants were instructed to name pictures of objects and animals in Spanish to the best of their knowledge, without consulting a dictionary and while alone in a quiet room. They were instructed to say “I don’t know” (in German) if a Spanish label for a picture was unknown. Four practice trials were administered after which participants could start the main part of the experiment via a button press.

Each trial started with the presentation of a picture in the centre of the screen. The audio recording started automatically at picture onset. Participants had a maximum of one minute for their response. The recording stopped either after one minute or when participants clicked a button to indicate that they were done

speaking. The recording was uploaded to the server, after which participants pressed a button to proceed to the next trial. The order of presentation of the items was identical for all participants, but different between sessions T1, T2 and T3 (see the Appendix for session-specific item lists).

Accuracy scoring. For each oral response, we counted how many phonemes were produced correctly and incorrectly. We chose this fine-grained coding to take partially correct responses into account (e.g. *sella* instead of *sello*; at T2: 5% in total and 14% of errors; at T3: 5% in total and 12% of errors; see also de Vos et al., 2018; Mickan et al., 2020). Incorrect productions could be either insertions, deletions or substitutions (see Levenshtein, 1966). Table 2 exemplifies the scoring procedure for the *sella* example.

Sella would be counted as having three correct phonemes and one incorrect phoneme. The vector of these two numbers (3,1) formed the basis for the dependent variable for statistical modelling. To provide descriptive statistics, we calculated an accuracy percentage (% correct for *sella*: $\frac{3}{4} = 75\%$). When participants corrected themselves, or otherwise needed multiple attempts to name a picture, the last utterance was scored. Synonyms were counted as correct productions and their phoneme count was adjusted as if the synonym was the target.

Predictor variables

Fluency tasks. *Material:* At each session, participants completed three German and three English fluency tests. In the interest of time and to avoid repeated measurements across sessions, we administered one letter fluency and two category fluency tests per language and session (i.e. six subtasks instead of three letter and three category tasks per language). The letters and categories differed across sessions and languages (see Table 3 and Supplementary Materials, S4, for details on selection procedure). All fluency tests were pre-tested with seven participants in German and English to ensure that they were at an appropriate level for speakers with low to moderate English proficiency.

Procedure: The fluency tests were administered in a fixed order that was identical for all participants: the English letter fluency task was followed by the two English category fluency tasks, which were followed in turn by the German letter fluency test and finally the

Table 2. Scoring example, phonetically transcribed.

Target word	s	e	ʃ	o
Participant’s production	s	e	ʃ	A
Scoring	correct	correct	correct	Incorrect (substitution)

Table 3. Letters and categories chosen for the fluency tests.

	English		German	
	Letter	Categories	Letter	Categories
T1	F	land animals, professions	P	kitchen utensils, vegetables
T2	A	clothes, fruit	M	hygiene/bathroom supplies, transportation means
T3	D	body parts, electronic devices	B	sports, office supplies

two German category tests. The letter and semantic category to be produced differed across sessions to avoid repetition or practice effects (see Table 3); however, the same categories were tested in each session for each participant to make differences in performance across participants informative. The fluency test also started with a microphone test, followed by the instruction to name as many words as possible, but to avoid proper names, compounds with the same head (e.g. fish, fishnet), and inflections or derivations of words (e.g. run, ran, running, or actor, actors, actress). Each test started with a countdown of five seconds during which the participant saw the category or letter for the upcoming task accompanied by a British or German flag indicating the response language, after which the recording started automatically. Following the standard for such fluency tests (e.g. Bolla et al., 1990; Gladsjo et al., 1999; Rosselli et al., 2002; Shao et al., 2014), the allowed response time was one minute, after which a pause screen appeared. Participants could start the next fluency task by clicking a button.

Answer Scoring: Answers were coded offline. Each existing, unique word that fulfilled the restrictions described in the instructions was counted as a valid word. The number of those valid responses was used as the score for each task. Due to technical difficulties, the last fluency task in each session failed to completely upload to the server for some participants, such that recordings for 20% of participants lasted only 20–50 s rather than a full minute. Rather than excluding these participants, we therefore decided to omit the last fluency task. This means that for the German category fluency score, we only have one category (always the first listed per session in Table 3), rather than two.

To arrive at a fluency change score, we subtracted the number of valid words for each given letter or category at T2 from their respective T3 scores and divided this number by the T2 score, thus indicating how much a participant's fluency increased or decreased from T2 to T3. For the English category fluency, we averaged over the two category scores; all other scores reflect performance on just one task per session. Because different

categories were used in each session, mean fluency differences between T2 and T3 for each language are not interpretable but changes in fluency by participant across sessions can indicate a trade-off in accessibility between languages.

Questionnaires. At each session, participants filled in a questionnaire asking for frequency of use ratings in all languages, proficiency self-ratings in Spanish and English, their motivation to learn Spanish, and a few session-specific questions about their study abroad and their time back in Germany. The full list of questions for each session in their original order can be inspected in the supplementary materials (S5). Below we will only describe those parts of the questionnaires that we included in the analyses reported below. Other questions were either not suited for analysis (open questions), not relevant for forgetting rates, or were measured for outlier detection rather than analysis.

Frequency of use ratings: At each session and once every month in between, we asked participants to estimate their current frequency of use in Spanish, German, English and other languages for the following four domains: reading, writing, speaking, and listening. Estimates were given in percentages, the sum of which had to add up to 100% in each domain (e.g. Spanish 50%, German 20%, English 30%), such that a reported percentage reflected the relative time someone spent, for example, reading in a particular language out of the total amount of time they spent reading. Note that this percentage does not take into account the total amount of time that someone spends reading, speaking, writing and listening. To get a sense of the latter, we additionally asked for estimates in hours (at T2 and T3 only, separately for a number of different contexts, see Supplementary Materials, S4). These additional estimates suggested that our participants did not differ much in the amount of time they spent doing particular activities (all $SDs < 1.6$ h), which reinforced our choice to use the percentage scores for analysis.

For analysis, we averaged over the four domains, resulting in one frequency of use percentage per language, session and participant. We chose to average over domains in an effort to reduce the number of predictors for analysis, and because correlations across domains at both T2 and T3 were high (all $r's > .60$; see Supplementary Materials, S7, for a complete correlation matrix). We further averaged across all measurements *after* T2 (that is all in-between measures and T3), such that the final frequency of use percentage for each language reflected the average amount of time a participant spent using that language after their study abroad. As we explain in the Introduction, we believe that this

average provides a more accurate estimate of the total average frequency of use over the attrition period than a retrospective estimate of frequency of use obtained at T3 only (as was often the case in previous studies).

Because the frequency of use indications for all languages together add up to 100%, they are mutually interdependent, and cannot be entered into a statistical model together. Because of this, we reduced the three frequency of use measures (for English, German, and Spanish) to two: Spanish frequency of use and the ratio of English over German use. The ratio reflects whether someone predominantly used German (L1) or English (L2) during the time not spent speaking Spanish. A ratio greater than 1 reflects more use of English compared to German, a ratio of 1 reflects equal use of the two languages, and a ratio between 0 and 1 reflects more use of German compared to English.

Motivation questionnaire: The motivation questions were identical in all sessions, and were taken in part from Mehotcheva (2010) and in part from Gardner's Attitude and Motivation Test Battery (AMTB, Gardner, 1985).¹ All 37 questions were translated into German, and were adjusted to the study abroad context when necessary (see Supplementary Materials, S5, for the full list of questions). Answers were given on a scale from 1 ("completely disagree") to 7 ("completely agree"). 25% of the questions were mirrored (with a low score meaning high motivation) and scores for those reversed before analysis. Questions were divided over 5 subcategories (in line with Gardner's taxonomy) asking participants about their:

- (1) General motivation to learn foreign languages (11 questions)
- (2) Attitude towards the Spanish people (9 questions)
- (3) Integrative motivation to learn Spanish (i.e. for social & intrinsic reasons) (6 questions)
- (4) Instrumental motivation to learn Spanish (i.e. for pragmatic/utilitarian reasons, such as finding a job) (5 questions)
- (5) Anxiety / nervousness related to using Spanish (6 questions; high score = high anxiety)

For modelling, we followed Gardner's suggestion to average each participants' scores on the first three subcategories to arrive at an overall score of integrative motivation per session. This choice was reinforced by the fact that scores on these three categories correlated highly positively with one another (all r 's at T2 and T3 $> .50$, see Supplementary Materials, S7, for the full correlation matrix), while scores on instrumental motivation and anxiety did not correlate strongly with any of the other categories (r 's $< .30$). We thus kept the latter two

subscales separate. Finally, because we were interested in how someone's *average* motivation to learn Spanish after the study abroad would affect their Spanish vocabulary development from T2 to T3, we chose to average T2 and T3 scores for each of the three subcategories for each participant.

Type of Spanish input: At both T2 and T3, we asked participants whether they were currently regularly speaking Spanish, and if so, to what percent they were doing so with Spanish native speakers as compared to non-native Spanish speakers. The total percentage had to add up to 100%. This question was only asked to participants who indicated that they were still actively using Spanish, which was 47 out of 97 participants at T3. Hence, we could not include this variable in the main analysis. We did, however, run an additional analysis including this variable on this subset of participants. For this secondary analysis, we used the average percentage of native input across T2 and T3 as predictor for Spanish forgetting rates.

Attrition self-judgment: Finally, at T3, we asked participants to judge whether their Spanish had improved or worsened since returning from Spain. Judgments were made on a scale from 1 (worsened a lot) to 7 (improved a lot), with 4 reflecting no (perceived) change. The resulting score was entered into the statistical models as is, in order to find out whether participants' perceived attrition aligns with our objective Spanish proficiency measure.

Doors test. At T3 only, we asked participants to complete the Doors test, a visual long-term memory test developed by Baddeley et al. (1994). Out of the 100 target-foil sets available on the "Doors of memory" website (<https://www.york.ac.uk/res/doors/resources.shtml>; Baddeley et al., 2016), we chose 30 sets to make the test short enough to administer online (see the Supplementary Materials, S8, for a list of doors). The test started with an encoding phase, in which participants saw a sequence of 30 target door pictures, each for one second, separated by blank screens of 200 ms. Participants were told to remember the doors and that they would later be tested on them. In the test phase, participants saw 30 picture assemblies, consisting each of one previously seen door (i.e. the target) and three foils, matched in style and colour to the target door. The participants' task was to select the previously seen target door by clicking on it. There was no time limit and participants did not get feedback on their selection. For analysis, we calculated the percentage of correctly recognised target doors.

Overview of predictors. All variables used as predictors of forgetting rates from T2 to T3 are summarised in Table 4. Next to the predictors described in detail above, we also

included amount of experience with Spanish prior to the study abroad and T2 performance as predictors (see Introduction for motivation), making for a total of 14 participant-level predictors. Finally, word frequency and cognate status were included as item-level predictors. Descriptive statistics about each predictor appear in the Supplementary Materials, S6.

Modelling

To investigate individual differences in forgetting rates, we ran logistic mixed effects models in R (Version 3.5.1, R Core Team, 2018), using the lme4 package (version 1.1–21, Bates et al., 2015) and the optimiser bobyqa. The dependent measure for all these models was the odds of correctly producing a phoneme for a given target word in the Spanish vocabulary test. A two-column matrix with the number of correct and incorrect phonemes for each target word at both T2 and T3 was passed to the model as dependent variable (see Accuracy coding above, this is one of multiple ways of specifying the response variable in binomial models, see <https://www.rdocumentation.org/packages/stats/versions/3.2.1/topics/family>; see also de Vos et al., 2018; Mickan et al., 2020, for examples of this approach).

Forgetting rates are reflected by a change in accuracy from T2 to T3. Hence, we included a variable Session (T2 vs T3) that was effects coded (−0.5, 0.5), such that a negative beta estimate for this variable reflects forgetting (i.e. a decrease in accuracy from T2 to T3) and a

positive estimate reflects learning (i.e. an increase in accuracy from T2 to T3).

To examine changes in forgetting rates, we entered each predictor in interaction with Session into the model. A significant interaction term means that a predictor affects the change in accuracy from T2 to T3: a positive estimate reflects that with every unit increase in the predictor variable, vocabulary size (the dependent variable) at T3 became larger relative to T2 (i.e. a learning effect). A negative interaction estimate, in turn, reflects *decreasing* T3 vocabulary scores compared to T2 (i.e. forgetting) with larger predictor values.

For each of the above listed predictors, we checked whether their inclusion in a model with Session significantly improved model fit compared to a baseline model with Session as the only predictor. If inclusion of a predictor indeed improved model fit, as assessed via chi-square model comparisons, the predictor was later included in the final full model, otherwise it was discarded (see Supplementary Materials, S9, for model comparison outcomes). In the final model, we then entered all significant predictors, each in interaction with the Session variable, together. We did this in order to reduce the number of predictors in the final model and to arrive at the most parsimonious final model justified by the data. All models, both the separate initial models as well as the final full model, included random intercepts for both Subject and Item, and all *p*-values were calculated by model comparison, using chi-square tests, omitting one factor at a time.

Table 4. Overview of predictor variables assessed via model comparison (German = L1, English = L2, Spanish = L3).

Participant-level predictors	
Predictor name	Description
Frequency of use	
Spanish	average frequency of use of Spanish (in %), averaged over the four domains (speaking, writing, listening, reading) for the entire period after returning to Germany (i.e. all measures taken after T2)
English-German ratio	ratio of English over German use, averaged over the four domains for the entire period after returning to Germany. score = 1 → equal use of German and English score < 1 → more German than English score > 1 → more English than German
Fluency	
English letter	relative difference in fluency between T2 and T3 ((T3-T2)/T2) on the English letter test
English category	relative difference in fluency between T2 and T3 ((T3-T2)/T2) on the average over both English category tests
German letter	relative difference in fluency between T2 and T3 ((T3-T2)/T2) on the German letter test
German category	relative difference in fluency between T2 and T3 ((T3-T2)/T2) on the first German category test
Motivation ²	
Integrative	average integrative motivation score at T2 and T3
Instrumental	average instrumental motivation score at T2 and T3
Anxiety	average anxiety score at T2 and T3
Memory capacity	% correct score on the Doors test
T2 Spanish proficiency	% correct in the Spanish vocabulary test at T2
Amount of experience with Spanish ³	amount of experience with Spanish prior to the study abroad (in years), as reported in Table 1
Attrition self-judgment	subjective rating of having improved or gotten worse in Spanish since returning from Spain, indicated on a 7-point Likert scale
Amount of native Spanish input	average % of native Spanish input (regardless of the total amount of Spanish input) across T2 and T3
Item-level predictors	
Word frequency	Spanish log frequency
Cognate status	distinguishing non-cognates from cognates (in any of the three languages, e.g. German-Spanish-English or English-Spanish)

Next to the main analysis, we also had three secondary research questions (see Introduction), which will be described in the Results section.

Results

Descriptive statistics for the dependent variable and the predictor variables

Spanish picture naming performance

Figure 3 shows participants' performance in the Spanish vocabulary test at T1, T2 and T3. On average, participants learned Spanish words while abroad (absolute learning rate between T1 and T2: $M = 17\%$, $SD = 8\%$, range = $-1\% - 42\%$) and forgot words after returning to Germany (absolute forgetting rate between T2 and T3: $M = 4\%$, $SD = 6\%$, range = $-12\% - 21\%$). However, as the by-participant data show (plotted in light- and dark-grey lines in the background), there is a lot of variation. Zooming in on T2 and T3, it turns out that while the majority of participants forgot some Spanish, some forgot much more than others, and some participants even improved from T2 to T3. It is these individual

differences that we hope to explain with the above-listed predictor variables.

Distributions of predictor variables

Frequency of use. Figure 4 shows how frequency of use changed throughout the duration of the study (panel A), as well as how the resulting two frequency of use predictors (for Spanish, and the ratio between English and German use) for the time after T2 are distributed (panel B). After leaving Spain, participants spoke German the vast majority of the time and very little Spanish and English. The rather narrow standard error around the mean in this plot, as well as the corresponding histogram in panel B, furthermore show that there is relatively little variability in the frequency of use measures, especially with regard to the ratio of English to German use. None of our participants used more English than German (values are never above 1) and the majority used far more German than English. For Spanish use, the distribution is also right-skewed and most participants used Spanish 10% of the time or less. **Fluency.** Participants' performance on the fluency tasks at T2 and T3, as well as

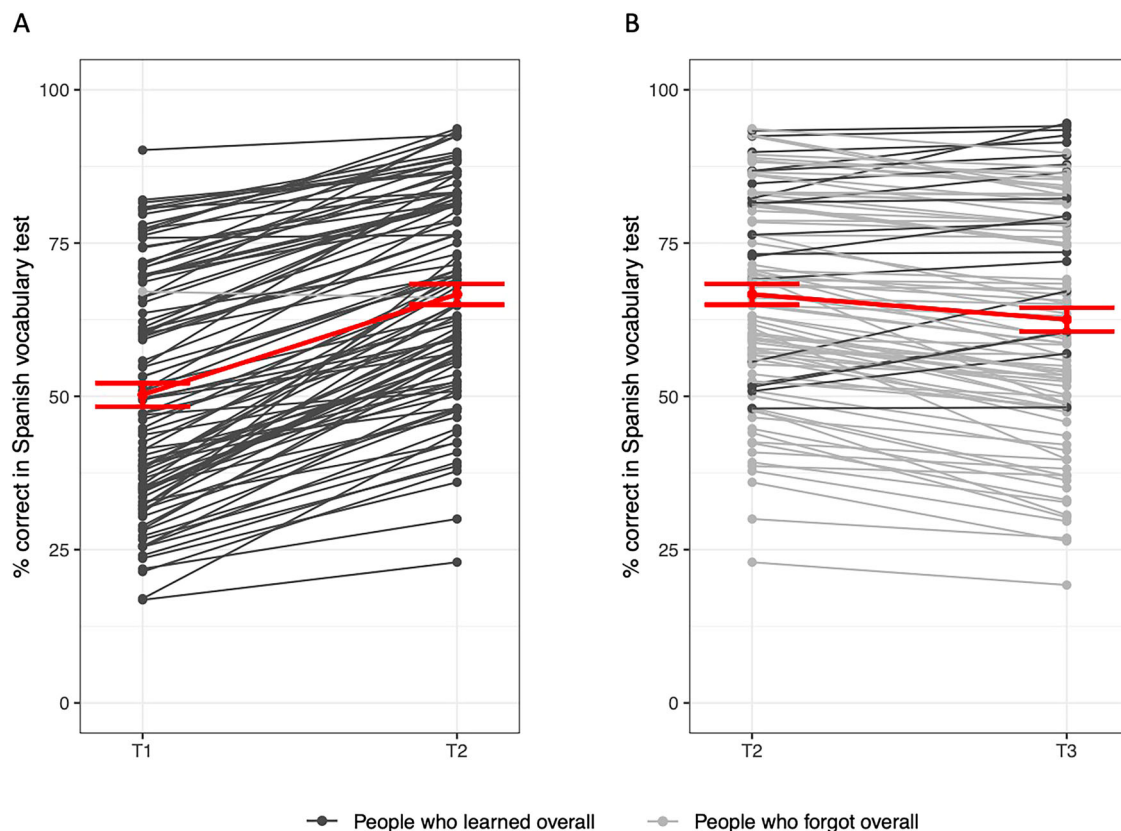


Figure 3. A. Participants' performance on the Spanish vocabulary test at T1 and T2. Dark grey lines represent people who learned on average while abroad, light grey lines represent people who forgot while abroad ($N = 1$). B. Participants' performance on the Spanish vocabulary test at T2 and T3. Dark grey lines reflect participants who learned on average after returning from abroad, light grey lines reflect participants who forgot after returning to Germany. In both subplots, the red line reflects the group mean and error bars correspond to the standard error around the mean.

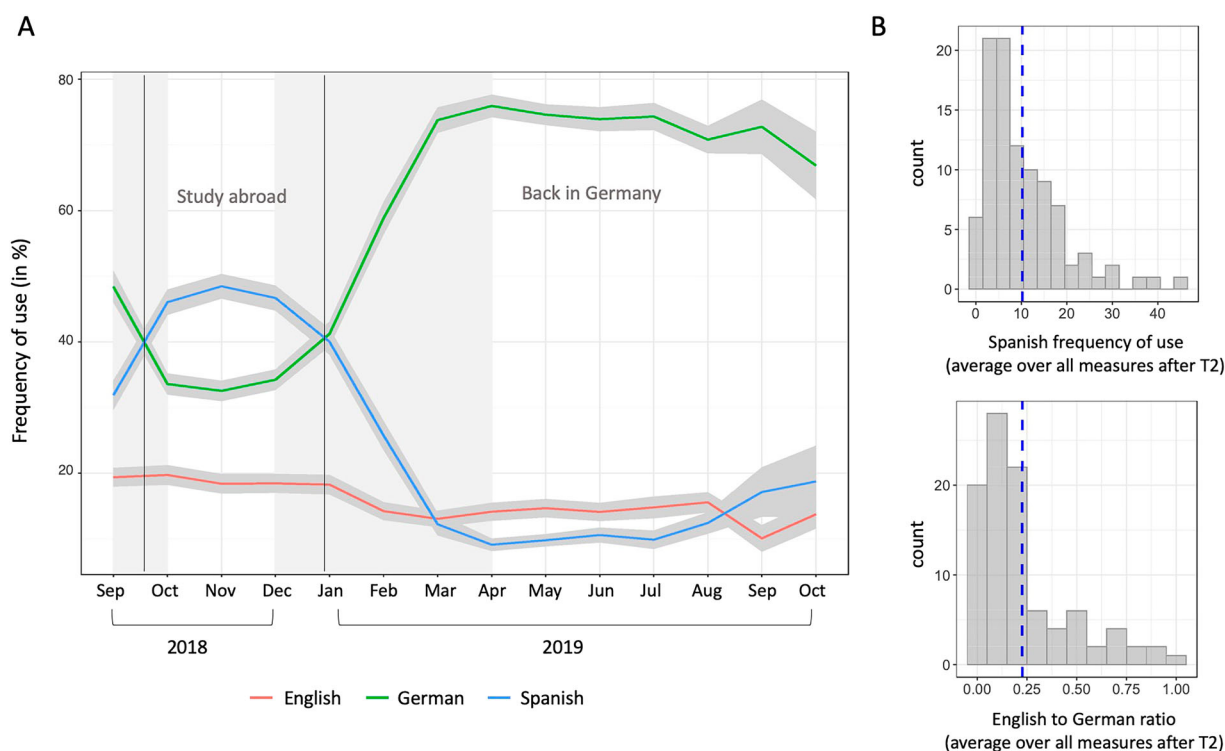


Figure 4. A. Average frequency of use for each language throughout the duration of the study abroad as well as their time back in Germany. Grey areas reflect the standard error around the mean. Vertical stripes indicate the average start and end date of the study abroad and grey areas around those averages reflect the absolute ranges of start and end dates respectively. B. Histograms for the two frequency of use predictors for modelling. The dashed blue line reflects the mean for each variable.

histograms for the resulting predictors can be inspected in Figure 5. Because the to-be-named categories were different at each session, absolute changes from T2 to T3 are not directly interpretable. We are instead interested in whether relatively large or small changes in (German and English) fluency scores from T2 to T3, compared with the other participants, predicted Spanish forgetting rates.

Motivation. Average motivation scores for each subcategory at T2 and T3, and the distributions of the corresponding predictor variables, can be inspected in Figure 6. Participants were, overall, very motivated to learn Spanish, and were so more out of personal interest and affinity with the language than for practical reasons (compare integrative with instrumental motivation). There was also much less variability in participants' integrative motivation compared to their instrumental motivation and their anxiety to speak Spanish. Moreover, on average, participants' motivation, both instrumental and integrative, as well as their anxiety to speak Spanish changed minimally from T2 to T3. Again, there was considerable variability between participants though. As explained above, for analysis, we averaged across T2 and T3 for each participant and each motivation subscore. In doing so, we approximated each participant's average motivation throughout the time period under investigation. As explained in

Table 4, only the instrumental motivation score entered the full model reported below. Including all three motivation scores was not possible with the current sample size, and we chose instrumental motivation because (as confirmed via model comparison) it was the best predictor of forgetting rates out of the three subscores.

Remaining Participant-Level Predictors. Distributions for the remaining predictors can be inspected in Figure 7. Panel A shows the memory performance measure: Performance on the Doors test was mostly above the chance level of 25% ($M = 53\%$, $SD = 15\%$). Panel B shows the average performance in the vocabulary test at T2: T2 performance varied ($M = 66\%$, $SD = 17\%$), but not a single participant was at floor or at ceiling (range = 23–94%). Panel C shows self-judgments of attrition: The average participant thought their Spanish had not changed after moving back to Germany ($M = 3.92$, $SD = 1.57$). Again though, there was considerable variation in participant's attrition self-judgments with answers spanning almost the full scale of answers (from 1 "got a lot worse" to 7 "improved a lot").

Between-predictor correlations. A correlation plot for all 13 participant-level predictors is shown in Figure 8. The strongest associations exist between Spanish frequency of use and T2 performance ($r = .51$) and

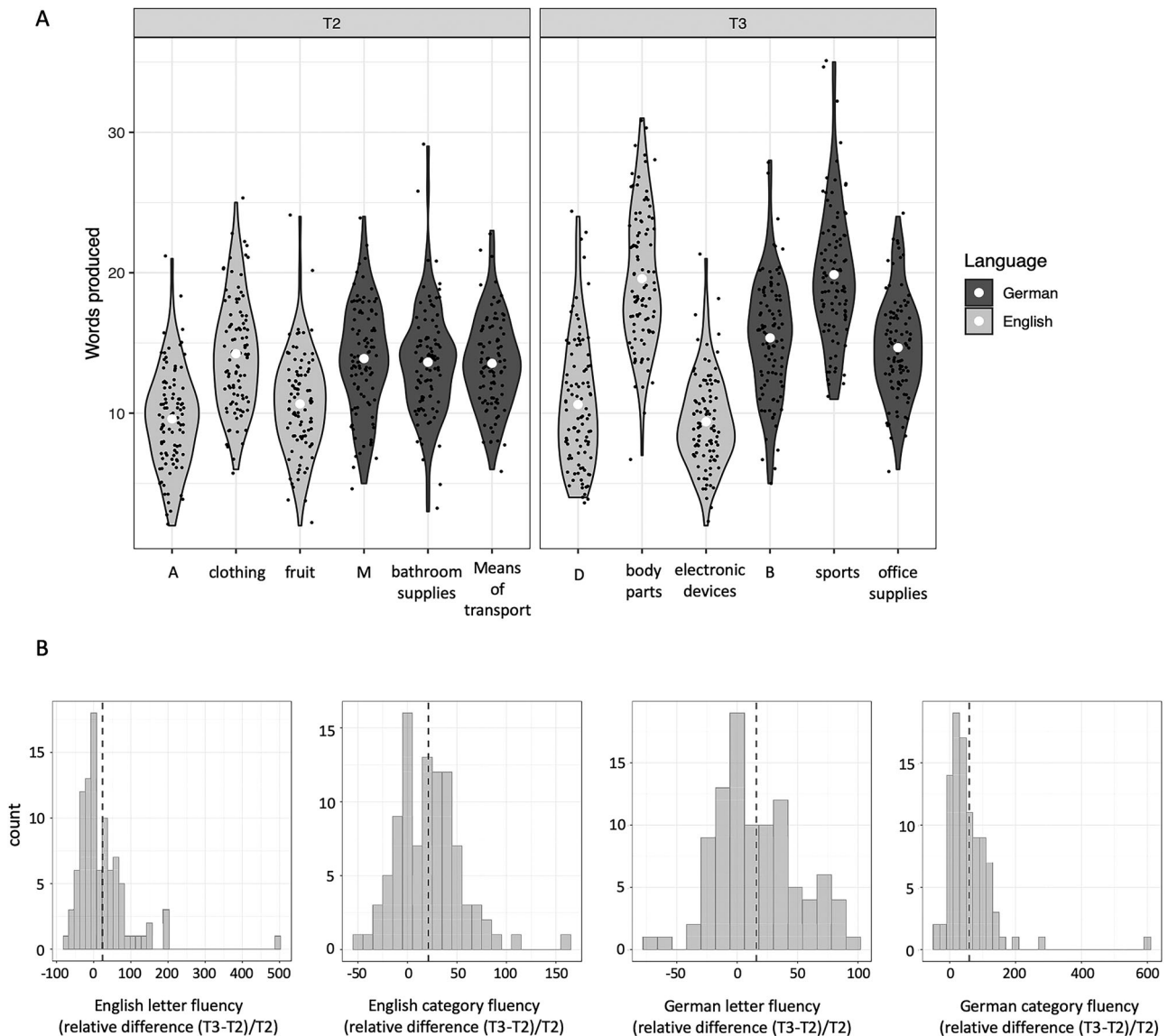


Figure 5. A. Violin plots for the number of words produced in each of the fluency tasks at T2 and T3. Categories/letters are plotted in the order that they were administered in. White dots represent means per task. Black dots represent individual participants. Violin plot outlines represent the distribution of the data. B. Histograms for the four resulting predictor variables used for modelling. Dashed blue lines reflect the mean for each predictor.

Spanish frequency of use and the motivation questionnaire scores: integrative ($r = .36$) and instrumental motivation to learn Spanish ($r = .49$) and anxiety to speak Spanish ($r = -.44$). T2 performance also correlated positively with integrative motivation ($r = .42$) and negatively with anxiety to use Spanish ($r = -.44$). No correlations exceeded .51, and all variance inflation coefficients are below 1.8, indicating that our predictors are sufficiently independent to be included within one explanatory model (Forthofer et al., 2007). Out of all 15 predictor variables (including the two item-level predictors), ten made it into the final model (see Supplementary Materials, S8, for the results of the model comparisons and Table 5 below for the final full model outcome).

For the subset of participants ($N = 47$) that we used for the analysis including the amount of native Spanish input predictor, the same relationships between predictors hold (see Supplementary Materials, S10, for the corresponding correlation matrix). The Spanish input predictor does not correlate highly with any of the other predictors (all r 's $< .22$).

Regression model outcomes

What predicts forgetting rates after a study abroad?

In the main analysis we asked whether the extent to which participants forgot Spanish after returning to their home

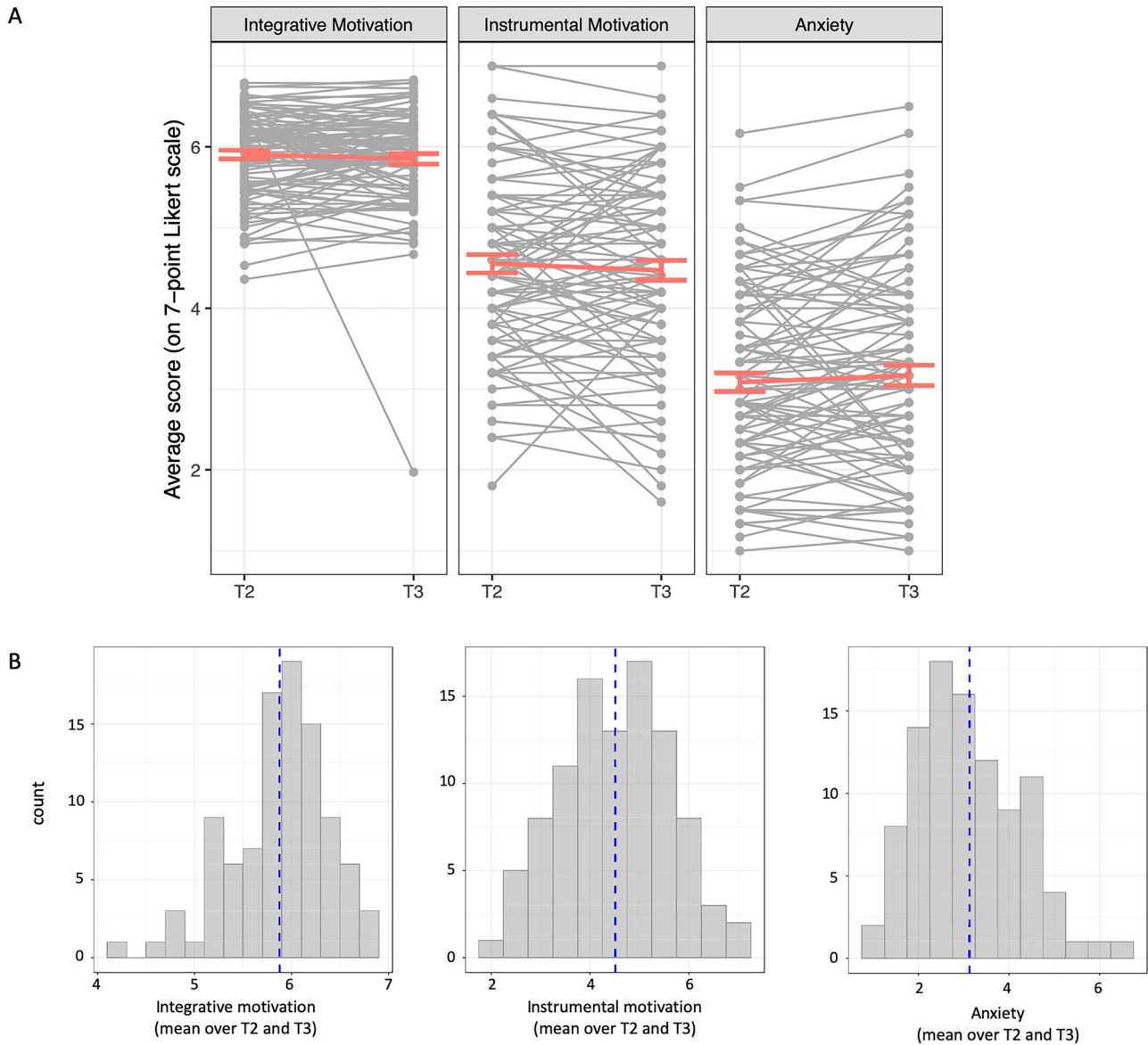


Figure 6. A. Average scores on each of the three subparts of the motivation questionnaire at T2 and T3. Light grey lines and dots reflect participant averages. Red lines reflect the means with the error bars denoting the standard error around the mean. B. Histograms for the three resulting predictor variables. The dashed blue lines reflect the mean for each predictor.

countries could be predicted by any of the participant- and/or item-level predictors discussed above. Model outcomes from the mixed effects logistic regression that we ran to answer these questions can be inspected in [Table 5](#).

We will first discuss the participant-level predictors, then the item-level predictors. In both cases, we will only discuss significant main effects if the predictor did not also significantly interact with session. For predictors that significantly modulated forgetting rates, their relationship to Spanish performance is illustrated in [Figure 9](#). Note that these plots reflect a simplified visualisation of the otherwise complex interactions between the predictors and the session variable (for the original

predictor plots from the mixed model, please consult the Supplementary Materials, S11).

First, we observed a main effect of **Session** such that participants made more errors in the Spanish vocabulary test at T3 than at T2. This means that after roughly half a year back in their home country, participants had overall forgotten some of the Spanish they used to know at the end of their study abroad ($M = 4\%$, $SD = 6\%$). This main forgetting effect was modulated by a number of factors, including, most importantly, **Spanish frequency of use**: people who still used Spanish relatively frequently after leaving Spain forgot less on average (and in fact even continued learning) compared to individuals who used less

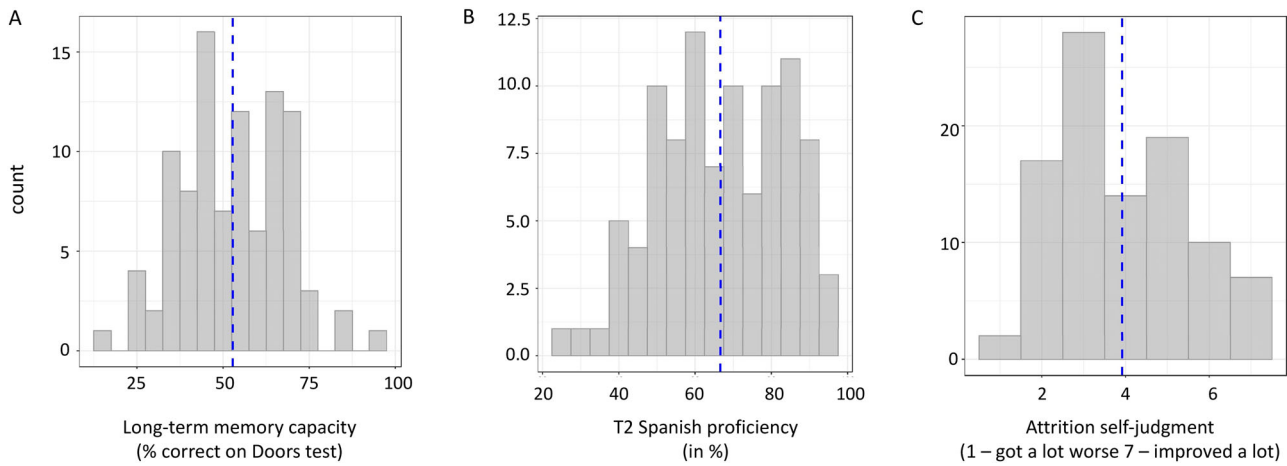


Figure 7. Histograms for the remaining predictor variables used for modelling. The dashed, blue lines reflect the mean for each predictor.

Spanish (see Figure 9 for visualisation). **English and German letter fluency** also significantly modulated forgetting rates, and did so in opposite ways (Figure 9). People whose English letter fluency scores (relative to other participants) increased from T2 to T3 forgot less Spanish than people whose English letter fluency

decreased.² For German letter fluency, we observed the opposite pattern: participants whose German letter fluency scores increased from T2 to T3 forgot more Spanish than people whose German letter fluency scores decreased. Possibly, the fact that we were able to use only one category per session instead of the more



Figure 8. Pearson correlation matrix for all 13 participant-level predictors from the main analysis. Colours indicate the strength of the correlation (Pearson’s r) with shades of blue indicating negative and shades of red indicating positive correlations. Predictors in bold made it into the final model.

Table 5. Logistic mixed effect model output for main analysis.

	Fixed effects			
	Estimate	SE	z	$p(\chi^2)$
Intercept	1.95	0.17	11.71	<.001
Session	-0.32	0.02	-15.79	<.001
<i>Participant-level predictors</i>				
Spanish frequency of use	0.10	0.03	3.08	.002
Amount of experience with Spanish	0.04	0.03	1.42	.156
Attrition self-judgment	0.05	0.03	1.64	.101
English category fluency	-0.02	0.03	-0.73	.467
English letter fluency	0.01	0.03	0.49	.623
German letter fluency	-0.02	0.03	-0.78	.434
Instrumental motivation	-0.01	0.03	-0.20	.842
T2 performance	1.34	0.03	45.93	<.001
Session * Spanish frequency of use	0.21	0.02	10.08	<.001
Session * Amount of Spanish experience	0.07	0.02	3.89	<.001
Session * Attrition self-judgment	0.08	0.02	4.50	<.001
Session * English category fluency	0.00	0.02	0.27	.789
Session * English letter fluency	0.07	0.02	4.74	<.001
Session * German letter fluency	-0.05	0.02	-2.93	.003
Session * Instrumental motivation	-0.01	0.02	-0.46	.643
Session * T2 performance	-0.05	0.02	-2.63	.009
<i>Item-level predictors</i>				
Word frequency	1.03	0.16	6.63	< .001
Cognate	2.19	0.33	6.63	< .001
Session * Word frequency	-0.03	0.02	-1.88	.060
Session * Cognate	0.12	0.04	3.04	.003
<i>Random effects</i>				
Item	Intercept	3.27	1.81	
Subject	Intercept	0.05	0.22	

Notes: Significant effects are marked in bold. SE = standard error; SD = standard deviation; $p(\chi^2)$ = Chi-square p -value; Var = variance.

typical three (see Methods) compromised the reliability of this measure, contributing to these unexpected results.

Next to frequency of use and fluency, participants with more Spanish experience before the study abroad forgot less than participants with less **prior Spanish experience**.³ Conversely, participants who performed better on the **vocabulary test at T2**, and who thus reached a higher Spanish proficiency level by the end of the study abroad, forgot more than participants with a lower recall score at T2. Finally, participants' own **judgment of attrition severity** was also predictive of observed forgetting rates: people who thought their Spanish got worse were the ones who indeed forgot the most (Figure 9). None of the other participant-level predictors significantly modulated forgetting.

On the item level, we observed a main effect of **word frequency**, such that successful recall in the Spanish vocabulary test, regardless of when it was administered (T2 vs. T3), was more likely for higher frequency items compared to lower frequency items. The interaction term between frequency and Session did not reach significance, but there was a numerical trend for high frequency items to be forgotten less often than low frequency items ($p = .060$). **Cognate status**, in turn, significantly predicted

forgetting, such that forgetting was more pronounced for non-cognates compared to cognates (see Figure 10).

Does studying abroad have long term (linguistic) benefits?

From the previous model, we learned that returning to one's home country resulted in attrition of foreign language vocabulary for the majority of people. A question that arises then is whether a study abroad has any long-term linguistic benefits at all, or whether vocabulary knowledge decreases to pre-study-abroad proficiency levels soon after leaving the foreign country. To answer this question, we ran another mixed effect logistic regression with the same random effects structure as the above models with data from all three time points (T1, T2 and T3) and with Session as the only fixed effect (dummy coded with T1 as reference level). The model outcome shows that the study abroad indeed had long-term benefits for our participants: participants learned while abroad (T2 performance exceeds T1 performance: $\beta = 1.40$, $z = 90.66$, $p \leq .001$) and forgot after moving back to Germany (see Figure 3). However, since their T2–T3 forgetting rates are smaller than their T1–T2 learning rates, performance at T3 was still significantly better than performance at T1 ($\beta = 1.042$, $z = 68.95$, $p \leq .001$).

Significant participant-level predictors of forgetting

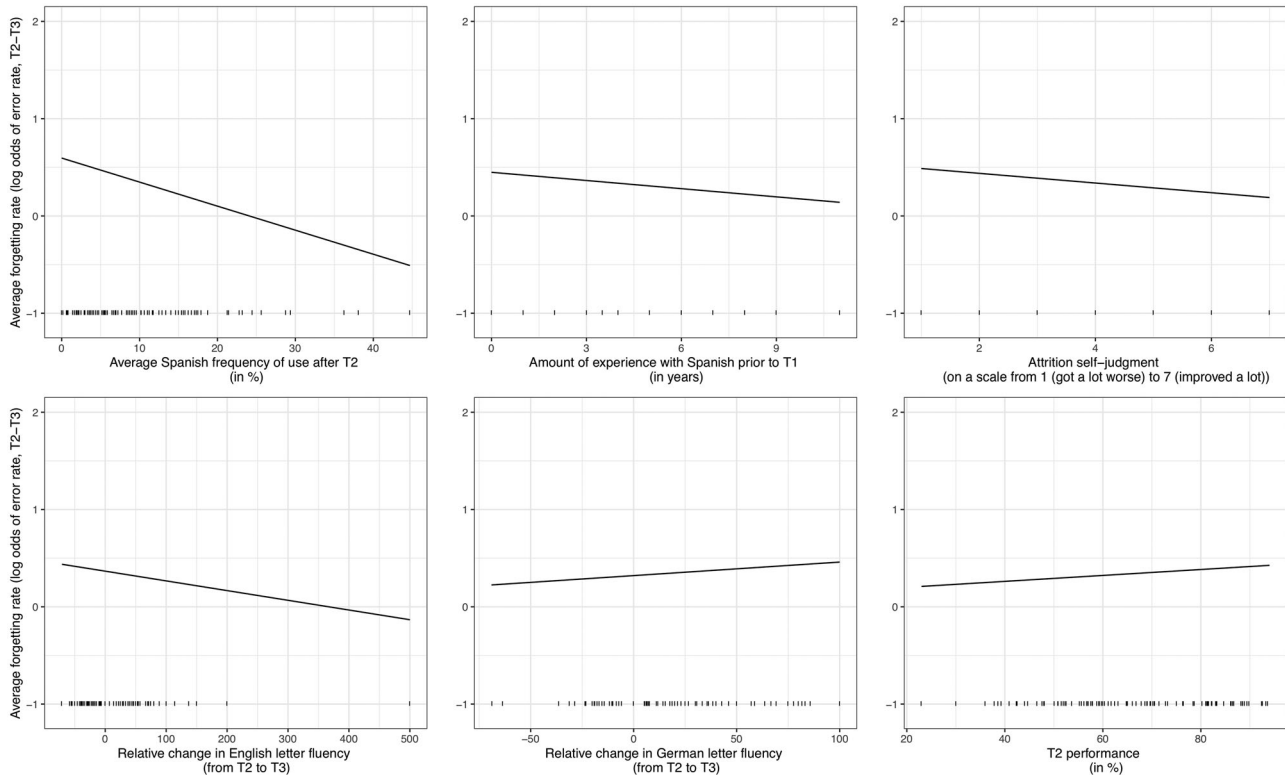


Figure 9. Participant-level predictor plots. The y-axis plots the average difference in error rates from T2 to T3 for each participant on a log odds scale, such that a positive difference reflects forgetting from T2 to T3, while a negative difference reflects learning from T2 to T3. The x-axes plot the respective participant-level predictors. Lines reflect the best-fit linear relationship between each predictor in the GLMs above, plotted as difference scores. Points (at bottom) represent scores on the predictor for individual participants.

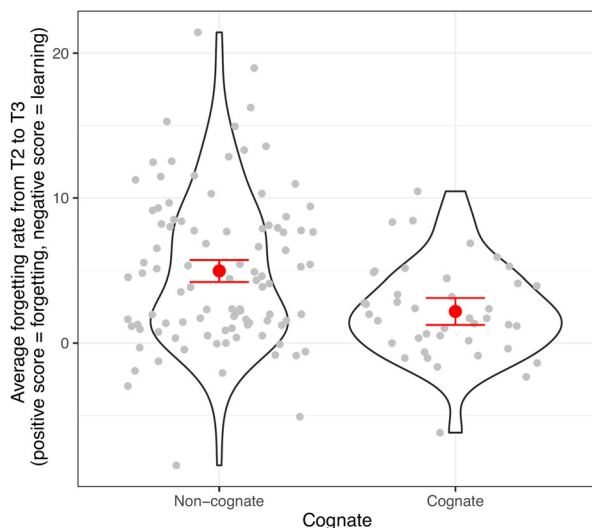


Figure 10. Violin plot of forgetting rates (T3 error % – T2 error %) for cognates and non-cognates separately, averaged over participants. Grey dots reflect items, red dots reflect the mean forgetting rate for cognates and non-cognates, respectively. Error bars reflect the standard error around the mean.

Testing the regression hypothesis

A long-standing debate in the attrition literature concerns whether forgetting mirrors acquisition, and hence whether what has been learned last is forgotten first. For our study, the Regression Hypothesis, as the former claim is also called, would predict that words learned between T1 and T2 (i.e. words not known at T1, but known at T2) have a higher probability of being forgotten at T3 than words learned *before* T1 (i.e. words known at both T1 and T2). To test this, we analyzed the data for words that were known at T2. For those words, we asked whether T3 Spanish performance was predicted by T1 performance. In modeling terms, this corresponds to running a mixed effect logistic regression on T3 Spanish performance with T1 performance as fixed effect (effects coded: $-0.5, 0.5$). This analysis showed that Spanish words unknown at T1 but known at T2 (learned while abroad) indeed had a higher probability of being forgotten than words that were already known at T1 ($\beta = -1.18, z = -30.99, p \leq .001$; mean accuracy at T3 for words

known at T1: 94%, $SD = 5\%$; mean accuracy at T3 for words unknown at T1: 73%, $SD = 15\%$).

Does the type of Spanish input matter for retention rates?

Finally, we asked whether the type of input matters for retention. Regardless of the time someone spends speaking Spanish, does someone who receives almost exclusively native input forget less than someone who receives more non-native, and hence potentially faulty Spanish input? This information was obtained for a subset of 47 participants out of the 97 participants, as explained above. We ran the same mixed-effect logistic regression model as in the main analysis on this subset with the average amount of native input at T2 and T3 in interaction with Session. The model outcome suggests that the amount of native input someone receives (regardless of the total amount of Spanish input) does predict forgetting rates: participants who received little to no native input forgot more than people who got a lot of native input, with the latter group showing evidence of learning rather than forgetting ($\beta = 0.06$, $z = 2.92$, $p = .004$). This result still obtains in a model including the seven significant predictors from the full model in the main analysis above. In this model, shown in Table 6, input type still explained a significant amount of the variance in forgetting rates. Moreover, except for cognate status and amount of experience with Spanish (prior to T1), all previously significant

predictors were still equally predictive of forgetting in this subset.

Discussion

The present study aimed at unravelling the driving forces behind foreign language attrition “in the wild”. How come we forget foreign language vocabulary, and what determines how fast and severe this lexical forgetting is? Based on recent lab experiments on FL attrition (e.g. Mickan et al., 2020), we hypothesised that language use would play a major role in determining the rate of attrition. More specifically, we assumed that continued target language use would positively impact FL retention, and conversely, that use of other languages would negatively influence FL proficiency. We also asked whether it matters *which* other language an attriter used the most (i.e. L1 vs. L2), and whether we can observe a trade-off in accessibility between those languages and the target foreign language.

In a large-scale longitudinal project, we followed a group of German learners of Spanish, who studied abroad in Spain for one semester. We evaluated their Spanish proficiency with a picture-naming vocabulary test at the beginning of the study abroad (T1), at the end of it (T2) and roughly six months post return to Germany (T3). Participants also completed fluency tests in German and English at each time point, as well as a questionnaire, asking, among other things, for current frequency of use indications and their

Table 6. Logistic mixed effect model output for analysis including amount of native input as predictor, run on subset of participants ($N = 47$).

Fixed effects	Estimate	SE	z	$p(\chi^2)$
Intercept	2.21	0.18	12.17	<.001
Session	-0.21	0.02	-9.04	<.001
<i>Participant-level predictors</i>				
Spanish frequency of use	0.10	0.04	2.49	.013
Amount of native input	0.03	0.04	0.88	.380
Attrition self-judgment	0.03	0.04	0.67	.504
English letter fluency	0.00	0.04	0.07	.943
German letter fluency	-0.04	0.04	-0.90	.371
Amount of Spanish experience	-0.01	0.04	-0.27	.788
T2 performance	1.07	0.04	27.00	<.001
Session * Spanish frequency of use	0.21	0.03	7.09	<.001
Session * Amount of native input	0.08	0.03	3.38	.001
Session * Attrition self-judgment	0.08	0.03	3.02	.003
Session * English letter fluency	0.06	0.02	2.54	.011
Session * German letter fluency	-0.06	0.03	-2.32	.021
Session * Amount of Spanish experience	0.03	0.03	1.09	.274
Session * T2 performance	-0.09	0.03	-3.50	.001
<i>Item-level predictors</i>				
Cognate status	1.00	0.18	5.43	<.001
Session * Cognate status	0.05	0.03	1.77	.077
<i>Random effects</i>				
	Groups	Var	SD	
Item	Intercept	4.57	2.14	
Subject	Intercept	0.05	0.21	

Notes: Significant effects are marked in bold. SE = standard error; SD = standard deviation; $p(\chi^2)$ = Chi-square p -value; Var = variance.

motivation to learn Spanish. With a logistic regression model, we then investigated which factors best predicted changes in Spanish vocabulary knowledge from T2 to T3. The longitudinal design made it possible to obtain fine-grained, participant-specific lexical attrition rates and hence enabled a much more precise analysis of the determinants of individual differences in FL lexical forgetting than would be possible with a cross-sectional study design.

Overall, we indeed observed forgetting after the first six months back in Germany: participants performed worse on the Spanish vocabulary test at T3 than at T2. Forgetting rates also varied considerably between individuals, with frequency of Spanish use between T2 and T3, quality of Spanish input (native vs. non-native), changes in German and English fluency, Spanish knowledge at the end of the study abroad, and years of Spanish experience prior to the study abroad predicting Spanish attrition rates.

The role of language use in foreign language attrition

Despite the undisputed role of language use in theories of language attrition (e.g. Köpke, 2002; Paradis, 2004), we are among the first to establish a clear relationship between target foreign language use and maintenance of FL skills in real attriters. Previous studies with real attriters have paradoxically often failed to observe a consistent relationship between the two (e.g. Bahrck, 1984a, 1984b; Mehotcheva, 2010; but see Huensch et al., 2019, and case studies by Ecke & Hall, 2013; and De Bot & Lowie, 2010, for demonstrations of a strong role for the use of an attriting FL in its accessibility). As we discussed in the Introduction, this failure may stem from the way in which language use was measured in those experiments. Given that many studies are cross-sectional rather than longitudinal in design, participants are often asked to estimate frequency of use once and in retrospect. In the current study, we averaged over multiple measures of frequency of use, taken once every month during the attrition period, asking for current rather than retrospective judgments. Moreover, instead of asking for ratings on a scale or for indications in hours and minutes, we chose a, possibly more straightforward and natural, percentage measure. We believe that these aspects combined resulted in a more accurate description of (average) Spanish frequency of use over the attrition period and hence are at least part of the reason why we were able to observe a clear-cut relationship between Spanish frequency of use and Spanish retention. We encourage future research to adopt similarly frequent, percentage-based frequency of use questionnaires.

Admittedly, by using percentages, we ignore information about the total amount of time that someone spends engaging with a language. One might argue that equating people who speak a lot with people who have few social contacts is problematic. For our specific population, this turned out not to be of any concern. Additional frequency of use indications in hours (at T2 and T3 only, separately for a number of different contexts, see Supplementary Materials, S4, for a list of questions) showed that our participants did not differ much in the amount of time they spent doing particular activities (all *SDs* <1.6 h). We think this is likely to be true for a lot of foreign language attriter populations, especially when the population is homogenous in terms of age, socio-economic status and cultural background, such as the population we recruited. Our study thus showed that among attriters comparable in terms of total absolute amount of language use, differences in the relative amount of use of different languages are a reliable predictor of forgetting rates.

Next to quantity of Spanish input, we also found that the *quality* of the input matters: participants who, regardless of the total amount of time they spent speaking Spanish, received mostly native input forgot less than participants who received less native and hence potentially faultier and less reliable input. In a model with both quantity and quality of exposure (as well as all other predictors), both factors appeared to be equally important. We are not aware of any other study that has explicitly tested whether the amount of native as compared to non-native input matters for foreign language retention. Nevertheless, the finding that the quality of the foreign language input matters resonates well with previous calls to account for the context of language use in studies on (both foreign and first) language attrition (Schmid, 2007, 2019).

Because the amount of Spanish language use is inversely related to the amount of use of all other languages combined (especially when using percentage measures), a positive relationship between Spanish retention and use entails a negative relationship between Spanish retention and the use of *other* languages. In that respect, our findings are in line with lab studies that report that speaking languages other than the target FL language hampers subsequent access to the FL (e.g. Levy et al., 2007; Mickan et al., 2020). It should be noted though that these lab studies' findings are not fully comparable to our results. In the lab, unlike in real life, researchers can keep target language use constant while manipulating non-target language use and can hence isolate the role of speaking other languages on target FL attrition (see Mickan et al., 2020). While doing this is impossible in real life, we still asked

whether it made a difference *which* language our participants spoke most during the time they did not speak Spanish. Mickan et al.'s (2020) lab study, and some more applied studies reviewed by Ecke (2015), suggest that other foreign languages interfere more with a foreign language than the mother tongue (the “L2 status effect”; Williams & Hammarberg, 1998). However, the English (L2) to German (L1) ratio variable that we introduced to answer this question did not reliably modulate forgetting rates, even in a separate model alone. While this means that we failed to replicate the lab findings from Mickan et al. (2020), we note that all of our participants used far more German than English (values are all below one and close to zero). Because there were no participants in our sample who used more English than German, we likely did not have enough variability to detect the effect that Mickan et al. (2020) observed in the lab. In real life, it might be difficult to find participants that use more English than German while living in Germany. An interesting alternative for future research might be to follow Germans who move to a country where they are immersed in English after a study abroad in Spain and see whether they suffer more from Spanish attrition than those that return to their L1 environment. For now, we can only conclude that for people who use their L1 a lot more than their L2, it does not matter how much more they use the L1, as L3 attrition rates were comparable regardless of the ratio of English to German use.

Finally, we also took fluency measures in English and German, hypothesising that relative fluency increases in both languages would predict proficiency decreases in Spanish. We found partial evidence for such a trade-off: participants whose German letter fluency scores increased the most relative to other participants, and hence participants who maintained their Germany fluency best (or in fact improved), were indeed more likely to forget Spanish vocabulary than those whose German letter fluency scores did not increase. Conversely though, and unexpectedly, relative increases in English letter fluency predicted increases, rather than decreases, in Spanish proficiency. From our data it thus appears that the two foreign languages, English and Spanish, co-develop and possibly even facilitate each other. Only the native language German shows the trade-off that we had expected based on previous lab research on interference-induced forgetting (e.g. Bailey & Newman, 2018; Mickan et al., 2020). While this is puzzling and appears to contradict Mickan et al. (2020), we note that our fluency scores may not capture the desired construct. Our trade-off hypothesis was based on the premise that fluency scores can be used as a proxy for

verbal ability in German and English, meaning that they should correlate positively with frequency of use in German and English respectively. However, the change in letter fluency from T2 to T3 in German and English did not correlate with average frequency of use between T2 and T3 in German ($r = .03$) and English ($r = .22$), nor did the change in English letter fluency correlate with our participants' perceived change in English proficiency, as assessed via self-ratings ($r = .1$). The same is true for the category fluency scores (all r 's $< .26$). The fluency scores thus do not appear to reflect what we hoped they would. Besides the already mentioned limitation that our one-time measurement of letter fluency per session (and language) may have been insufficiently reliable, this means that another possible explanation is that a large part of the variance in fluency scores could be caused by other factors, including executive control ability (as suggested by Luo et al., 2010; Shao et al., 2014) or the higher similarity of English and Spanish than German and Spanish.⁴ What is more, in the literature, category fluency scores have been much more reliably linked to verbal ability and vocabulary size than letter fluency scores (Shao et al., 2014). Thus, it is interesting that out of the two types of fluency scores, letter and not category fluency predicted changes in Spanish proficiency. All things considered, the fluency data should be taken with a grain of salt.

For future studies, a test of English and German proficiency (rather than a test of verbal ability) could make for a more straightforward test of the trade-off hypothesis. Should changes in L3 proficiency indeed correlate positively with changes in L2 and negatively with changes in L1 proficiency, it would mean that interference between two foreign languages is much less prevalent in real life (and when measured purely in naming accuracy) than the lab studies suggest, and that counter to what Mickan et al. (2020) concluded, L1 is the stronger interferer. To get a preliminary sense of whether our pattern was reliable or not, we ran a statistical model with the difference in English proficiency self-ratings (T3–T2) instead of English (letter and category) fluency performance. Here, we found a negative relationship between changes in (perceived) English and (observed) Spanish proficiency: in a model with only the change in English proficiency self-ratings and Session as predictors, participants whose English improved according to their own self-judgments forgot more Spanish than people whose English got worse ($\beta = -0.04$, $z = -2.24$, $p = .025$). This effect is in line with the trade-off hypothesis. The change in perceived English proficiency, however, no longer significantly predicted forgetting rates when included in the full model with all other predictors (excluding fluency scores, $\beta = -0.01$, $z = -0.41$,

$p = .683$), and note that we have no parallel measurements of self-reported German proficiency. Though not conclusive, these follow-up analyses cast further doubt on the fluency findings and the suitability of fluency tasks to approximate proficiency / verbal ability in a given language.

FL vocabulary knowledge prior to attrition onset (T2)

Participants who performed better in the vocabulary test at the end of their study abroad (T2) were more likely to forget phonemes from T2 to T3 than participants with poorer T2 performance. This finding appears to contradict earlier research which has shown that a higher level of FL proficiency prior to attrition onset is *beneficial* for foreign language maintenance (e.g. Bahrck, 1984a, 1984b; Mehotcheva, 2010; Murtagh, 2003; Weltens, 1988), or has no effect on maintenance (Engstler, 2012). A closer look at those studies, however, reveals that most of them are cross-sectional studies that lack participant-specific baselines. Some of them, in fact, do not assess the effect that prior foreign language proficiency has on forgetting rates, but rather just the effect it has on performance at a single measurement of attrition (i.e. performance at T3 alone rather than the change in performance from T2 to T3; e.g. Mehotcheva, 2010; Murtagh, 2003). That higher initial proficiency predicts better performance at a later testing point is not that surprising and is also the case in our data (see Supplementary Materials, S11, lower middle panel; the pink line has a generally positive slope), but this does not say anything about the amount of forgetting (i.e. change in knowledge) since the start of the attrition period.

Another difference between our study and previous research is how we defined “initial proficiency”. Most previous studies used proficiency self-ratings (Engstler, 2012; Mehotcheva, 2010; Murtagh, 2003), the number and level of courses taken in the FL prior to attrition onset, and past course grades (or a combination thereof; Bahrck, 1984a, 1984b; Weltens, 1988) as estimates of initial proficiency. Our T2 performance measure is a much more specific and objective measure of past FL vocabulary knowledge and given that it reflects prior performance on the same task, it is directly comparable to our attrition (T3) measurement. To the best of our knowledge, we are the first to show specifically that pre-attrition vocabulary size influences lexical forgetting rates.

Nevertheless, it might seem puzzling that we observe a *negative* effect of T2 performance on subsequent Spanish vocabulary retention rates. That someone who

knew more words is more likely to forget parts (i.e. phonemes) of those words over time, however, might just be a reflection of the fact that they had more to lose in the first place. One might wonder then whether participants who knew more at T2 forgot more only in *absolute* terms (the number of forgotten phonemes, as our model suggests), or also in *relative* terms (the percentage of forgotten words out of all words known at T2 baseline); more absolute forgetting could be less relative forgetting if sufficiently many words were known at the baseline. To answer this specific question, we ran a linear model on participants’ relative forgetting rates ($(T3-T2)/T2$) with all the participant-level predictors from the main mixed model (excluding item-level predictors and random effects per participant or item, which are not possible in this simplified model).⁵ Partially in line with what Bahrck (1984a, 1984b) and Weltens (1988) reported,⁶ this model revealed that participants with high T2 scores forgot *less* in relative terms compared to people with lower T2 scores. In summary, we thus find that high T2 performers forget *more words/phonemes in absolute terms, yet fewer words/phonemes in relative terms* (i.e. a smaller percentage of their original knowledge) than low T2 performers, which means that the retention rate of the former group was ultimately better.

Amount of experience with the foreign language

The amount of experience with Spanish prior to the study abroad also predicted forgetting rates in the main analysis. In line with previous research, participants with more years of Spanish experience were less likely to forget. In the Introduction, we discussed how including the amount of FL experience might appear highly correlated with T2 performance and thus redundant. The two variables, however, correlated with one another only moderately in our sample ($r = .21$), and consequently predicted different aspects of the variance in forgetting rates. As already mentioned, it is very possible that some participants are faster learners than others and hence reach the same level of proficiency in less time. What is more, the number of years of exposure is not necessarily indicative of the quality or even quantity of exposure to the language in those years. Interestingly, in an additional analysis with the amount of native language input during the attrition period as additional predictor, the amount of experience with Spanish prior to the study abroad was no longer predictive of forgetting, while T2 proficiency and frequency of use of Spanish during the attrition period continued to explain a large part of the variance. This pattern suggests, quite intuitively, that amount and quality of

recent language use are more important for successful retention than total amount of time spent learning a foreign language. This pattern emerged from a model with only 47 out of the total 97 participants though for whom these quality of input data were available, and hence needs replication before firm conclusions can be drawn.

Finally, it should be noted that unlike Mehotcheva (2010), length of the study abroad was not a significant predictor of forgetting rates and thus did not enter the final model. Study abroad length was significantly worse at explaining forgetting rates than prior amount of experience with Spanish. This is not surprising given that there was little variability in the length of the study abroad between our participants, compared to quite large variability regarding the length of prior experience with Spanish.

Motivation

Based on Gardner's Attitude and Motivation test battery, we asked participants about their integrative motivation (i.e. intrinsic desire to learn Spanish for social reasons), as well as their instrumental motivation (i.e. desire to learn Spanish for practical reasons) and their anxiety to speak Spanish. In separate models, all three motivational variables significantly modulated forgetting, such that higher integrative and instrumental motivation and lower anxiety to speak Spanish were beneficial for Spanish retention. Out of those three scores, only instrumental motivation entered the final model. Including all three scores was not possible with the current sample size, and instrumental motivation scores predicted forgetting rates slightly better than the other two in those separate models.

Though highly predictive of forgetting rates when included in a model on its own, instrumental motivation no longer predicted forgetting rates once frequency of use and other factors were also accounted for. The most likely reason for this is the relatively high correlation between Spanish frequency of use and instrumental motivation ($r = .49$): participants who were highly motivated to learn and maintain Spanish likely sought out more opportunities to speak it. By virtue of being correlated, the two variables explain partially the same variance; frequency of use, however, appears to do so better. This pattern demonstrates once again that all variables need to be considered together in one parsimonious model in order not to overestimate the contribution of single predictors. That we do not observe a beneficial effect of motivation in our main model is in line with previous failures to establish such a link

consistently (e.g. Mehotcheva, 2010; Xu, 2010; though see Wang, 2010).

Non-verbal memory capacity

Non-verbal long-term memory capacity did not predict forgetting of Spanish vocabulary, either in the full model or in a separate model by itself. To our knowledge, we are the first to assess the relationship between non-verbal and verbal long-term memory in a population of foreign language attriters. Our results may be taken to suggest that the two do not interact and that ability in one domain does not predict ability in the other. However, this conclusion is premature. That the memories that are encoded in the Doors test are episodic rather than semantic in nature (like foreign language vocabulary after consolidation) and the fact that the test ended in a receptive rather than a productive recall test might explain why the Doors test had such poor explanatory power in the model on productive Spanish vocabulary forgetting. Future studies might want to consider a test of productive non-verbal memory instead, which within the scope of this online experiment, however, was unfortunately not possible to implement. Another possibility is that participants did not pay enough attention during the encoding phase and that their performance on the Doors test thus does not reflect their true non-verbal long-term memory capacity. Even though participants performed above the chance level of 25% on average, performance overall was still rather low ($M = 50\%$). The online set-up and the fact that the encoding phase did not require responses from participants and was automatically paced possibly encouraged participants to take the test less seriously. Future research should improve upon those aspects. Either way, a more complete assessment of non-verbal long-term memory capacity is necessary before drawing firm conclusions on its relation to verbal memory capacity.

Attrition self-judgments

Participants who thought their Spanish decreased indeed actually forgot more than participants who thought their Spanish had not changed (or had even improved). Participants who thought they improved, however, on average still performed worse on the Spanish vocabulary test at T3 than T2, suggesting that, unlike in other studies before (e.g. Weltens, 1988), our participants underestimated, rather than overestimated, their attrition. This underestimation might be related to the fact that participants were asked to judge their proficiency overall, rather than their vocabulary knowledge

specifically. Given that some of the words we asked for in the vocabulary test were low-frequency words and given that the attrition self-judgment was provided prior to the vocabulary test, our participants might not have been aware of their vocabulary gaps and hence rated their overall proficiency decline as less severe than the vocabulary test suggests.

Item level predictors: cognate status and word frequency

Finally, on the item level, we partially confirmed Weltens' conclusion (1988): in the main statistical model with all participants and predictors included, cognates were retained better than non-cognates. For word frequency, we also only found partial evidence for an effect on forgetting rates: in a separate model with only Spanish word frequency, low frequency words were more likely to be forgotten than high frequency words. This effect, however, was no longer statistically robust in the final model with all other predictors combined. Unlike Mehotcheva's (2010) results, our study thus suggests that word frequency plays only a minor role in vocabulary retention.

Does forgetting follow the reverse order of acquisition?

Next to investigating individual differences in foreign language attrition, we also asked whether we could find evidence for the regression hypothesis (RH) which, in its original formulation, states that forgetting follows the reverse order of acquisition (Jakobson, 1941) and hence that recently learned words are forgotten faster than words learned long ago. The longitudinal design of our study and the fact that we have a pre-study abroad baseline in addition to the pre-attrition baseline provided a unique opportunity to test this. Specifically, we asked whether words participants learned during their study abroad (known at T2 but not T1) were more likely to be forgotten than words they already knew before the study abroad (known at both T1 and T2). Our data indeed suggest that this is the case, providing robust evidence for regression in foreign language lexical attrition.

Some researchers have proposed that it is not the information learned last, but rather the information learned *least well* that is forgotten first (Hedgcock, 1991). In the current dataset, we have no measure of "how well" words were learned, and hence no way of testing this hypothesis. Future research might want to take a look at reaction times in picture naming as an estimate of how well the words were known at T2. If quality of learning matters more, naming latencies at T2 should

be a better predictor of T3 performance than order of acquisition.

A final note on overall attrition rates and study design

Regardless of the individual differences discussed so far, it is worth noting that observing significant attrition effects after just six months is quite remarkable, especially in light of the fact that some previous studies failed to observe any attrition after much longer periods of disuse (e.g. Engstler, 2012; Murtagh, 2003; Weltens, 1988). That we do observe forgetting in this rather short attrition period is probably in part due to the fact that we tested productive vocabulary recall, rather than receptive FL skills, as was the case in Weltens (1988). Much to the reassurance of all language learners out there, forgetting after the study abroad was overall still rather small (4% on average) and much less pronounced than learning during the study abroad (17%), meaning that the study abroad did ultimately result in long-term linguistic gains despite ensuing attrition.

Long-term in this case, of course, refers to only six months. We know from Bahrick's (1984a, 1984b) research that FL language skills decline steadily for the first three to six years, suggesting that the 4% loss that we observed on average after six months may represent just the beginning of the attrition process, and that we would have likely observed much larger forgetting rates had we tested our participants a few years after the study abroad. From the range of forgetting scores that we observed (−12% to 21%), it also becomes clear that not everyone forgot: some in fact continued learning (16 out of 97 participants). Participants who still used Spanish regularly, with friends in Spain or at university as part of their studies, were more likely to improve from T2 to T3 rather than to attrite. What we call the *attrition period* throughout this paper is hence only an attrition period in the aggregate. What is more, some of our participants, and in particular those that were tested later and hence had a somewhat longer time between T2 and T3, returned to Spain for their summer vacation. The longer the T2–T3 interval in our study, the higher the chance for re-exposure to Spanish, which is opposite to what attrition length usually should denote. We therefore refrained from including this variable in our statistical analyses. Amount of exposure to Spanish was already (and much better so) captured by our regular frequency of use questionnaires.

As a final note, we would like to highlight the online set-up of this study. When testing participants in person, following a large enough group of attriters over a long period of time is very time-consuming.

Online testing offers a convenient way out. Participant drop-out rates will still be high, but the automaticity of online testing and the fact that one is no longer constrained to one geographical location makes recruitment much easier and facilitates the collection of much bigger samples than possible with in-person testing. Naturally, online testing comes with its own set of drawbacks (e.g. technical difficulties, constraints on the types of tasks that can be administered and lack of control over how seriously participants take the tasks), but overall, we believe that the benefits of online testing – specifically the opportunity to do truly large scale longitudinal studies on FL attrition – clearly outweigh its costs.

Conclusion

The present study investigated individual differences in foreign language attrition as it unfolds within the first months after an immersive study abroad. We followed German university students throughout their study abroad in Spain, as well as throughout their first roughly six months back in Germany. In line with expectations, yet counter to most previous research on foreign language attrition “in the wild”, variation in forgetting rates, to a large part, turned out to be due to differences in the quantity and quality of language use: more frequent Spanish language use during the attrition period was clearly beneficial for Spanish vocabulary retention, and so was native (as compared to non-native) Spanish input, regardless of the total amount of input they received. Conversely, and partially in support of recent lab-based studies of FL attrition, increases in L1 German and, unexpectedly, decreases in L2 English verbal ability appeared to be detrimental for FL retention, at least in an environment where L1 use is predominant and where there is little room for use of other languages. Next to language use, Spanish vocabulary knowledge prior to attrition onset affected retention rates: participants who knew more words on average at T2 forgot absolutely more words (and/or phonemes in those words), yet in relative terms they forgot less than participants who knew fewer words at T2. Motivation to learn the foreign language and non-verbal memory capacity, in turn, had no influence on Spanish maintenance. Overall, the present study thus provides empirical evidence for the importance of continued language use for FL maintenance in real attriters. Moreover, using a longitudinal design and state-of-the-art statistical analyses, we were able to shed light on the complex interplay between language use and other determinants of FL attrition.

Notes

1. The questionnaire used by Mehotcheva (2010) was also based on the AMTB, yet was missing questions from one of Gardner’s subscales (anxiety).
2. A closer look at the predictor plot shows that there is an extreme outlier in the English letter fluency task, which might be contributing disproportionately to the effect. Running the model again without this outlier, however, shows identical results.
3. Amount of experience with Spanish was a significantly better predictor of forgetting rates than study abroad length ($p < .001$).
4. In line with a suggestion from a reviewer, one other factor that might have been at play here is the higher lexical similarity between English and Spanish, as also reflected in the number of cognates (English-Spanish: 41; German-(English-)Spanish: 18). Higher degrees of cross-lexical similarity may bring about co-development rather than competitive relationships between words of different languages.
5. In the linear model on relative forgetting rates, the data is reduced to just one value per participant and a lot of valuable information gets lost, which comes at the cost of accuracy and sensitivity in estimating fixed effects. With T2 performance being equal, forgetting absolutely less also corresponds to forgetting relatively less (a smaller proportion). For the assessment of all predictors other than T2 performance itself, our main model thus already implicitly takes initial T2 performance into account.
6. Bahrck (1984a, 1984b) and Weltens (1988) found that participants with different levels of initial training forget the same amount in absolute terms, which however corresponds to less forgetting in relative terms (a smaller proportion of their original knowledge). While we also find less forgetting in relative terms, our results differ from these previous studies in that we report more forgetting rather than comparable forgetting in absolute terms.

Acknowledgements

We would like to thank Fabian Schneider for help with the fluency test transcriptions and Wilbert van Ham and Arvind Datdien for implementing the picture naming and fluency tasks in Javascript.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Max-Planck-Gesellschaft: [grant number IMPRS fellowship 2016-2020].

Data availability statement

The data that support the findings of this paper are available at: <https://doi.org/10.34973/g8z6-vw16>.

ORCID

Anne Mickan  <http://orcid.org/0000-0002-2348-8921>
 James M. McQueen  <http://orcid.org/0000-0003-3734-6286>
 Laurel Brehm  <http://orcid.org/0000-0003-0424-8735>
 Kristin Lemhöfer  <http://orcid.org/0000-0003-0430-3968>

References

- Abbasian, R., & Khajavi, Y. (2010). Lexical attrition of general and special English words after years of non-exposure: The case of Iranian teachers. *English Language Teaching*, 3(3), 47–53. doi:10.5539/elt.v3n3p47
- Alharthi, T., & Al Fraidan, A. (2016). Language use and lexical attrition: Do they change over time? *British Journal of English Linguistics*, 4(1), 50–63.
- Baddeley, A. D., Emslie, H., & Nimmo-Smith, I. (1994). *The doors and people test: A test of visual and verbal recall and recognition*. Thames Valley Test Company.
- Baddeley, A. D., Hitch, G. J., Quinlan, P. T., Bowes, L., & Stone, R. (2016). Doors for memory: A searchable database. *Quarterly Journal of Experimental Psychology*, 69(11), 2111–2118. <https://doi.org/10.1080/17470218.2015.1087582>
- Bahrack, H. P. (1984a). Fifty years of second language attrition: Implications for programmatic research. *The Modern Language Journal*, 68(2), 105–118. doi:10.1111/j.1540-4781.1984.tb01551.x
- Bahrack, H. P. (1984b). Semantic memory content in permatore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, 113(1), 1–29. doi:10.1037/0096-3445.113.1.1
- Bailey, L., & Newman, A. J. (2018). Retrieval-induced forgetting and second language acquisition: Insights from a Welsh word-learning study. In J. M. Fawcett (Chair) (Ed.), Control processes in human memory: The role of retrieval suppression and retrieval practice. *Symposium conducted at the joint meeting of the Canadian Society for Brain, Behaviour and Cognitive Science and the Experimental Psychology Society, St. John's, CA*.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4 (Version 1.1–15) [R package]. <https://cran.r-project.org/web/packages/lme4/index.html>
- Bolla, K. I., Lindgren, K. N., Bonaccorsy, C., & Bleecker, M. L. (1990). Predictors of verbal fluency (FAS) in the healthy elderly. *Journal of Clinical Psychology*, 46(5), 623–628. doi:10.1002/1097-4679(199009)46:5<623::AID-JCLP2270460513>3.0.CO;2-C
- Brodeur, M. B., Dionne-Dostie, E., Montreuil, T., & Lepage, M. (2010). The bank of standardized stimuli (BOSS), a new set of 480 normative photos of objects to be used as visual stimuli in cognitive research. *PLoS ONE*, 5(5), e10773. <https://doi.org/10.1371/journal.pone.0010773>
- Brybaert, M., Buchmeier, M., Conrad, M., Jacobs, A. M., Bölte, J., & Böhl, A. (2011). The word frequency effect: A review of recent developments and implications for the choice of frequency estimates in German. *Experimental Psychology*, 58(5), 412–424. doi:10.1027/1618-3169/a000123
- Cohen, A. D. (1989). Attrition in the productive lexicon of two portuguese third language speakers. *Studies in Second Language Acquisition*, 11(2), 135–149. <http://doi.org/10.1017/S0272263100000577>
- Cuetos, F., Glez-Nosti, M., Barbón, A., & Brybaert, M. (2011). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicológica*, 33(2), 133–143.
- De Bot, K., & Lowie, W. (2010). On the stability of representation in the multilingual lexicon. In M. Pütz & L. Sicola (Eds.), *Cognitive processing in second language acquisition* (pp. 117–134). John Benjamins. <https://doi.org/10.1075/cecr.13.11bot>
- de Vos, J. F., Schriefers, H., & Lemhöfer, K. (2018). Noticing vocabulary holes aids incidental second language word learning: An experimental study. *Bilingualism: Language and Cognition*, 22(3), 500–515. <https://doi.org/10.1017/S1366728918000019>
- Ecke, P. (2004). Language attrition and theories of forgetting: A cross-disciplinary review. *International Journal of Bilingualism*, 8(3), 321–354. <https://doi.org/10.1177/13670069040080030901>
- Ecke, P. (2015). Parasitic vocabulary acquisition, cross-linguistic influence, and lexical retrieval in multilinguals. *Bilingualism: Language and Cognition*, 18(2), 145–162. <https://doi.org/10.1017/S1366728913000722>
- Ecke, P., & Hall, C. J. (2013). Tracking tip-of-the-tongue states in a multilingual speaker: Evidence of attrition or instability in lexical systems? *International Journal of Bilingualism*, 17(6), 734–751. <https://doi.org/10.1177/1367006912454623>
- Engstler, C. (2012). *Language retention and improvement after a study abroad experience* [Doctoral dissertation], Northwestern University. Retrieved June 3, 2020, from <https://www.linguistics.northwestern.edu/documents/dissertations/linguistics-research-graduate-dissertations-engstlerdissertation2012.pdf>
- European Commission. (2020). *Erasmus+ annual report 2018*. <https://doi.org/10.2766/989852>
- Fernández, J., & Gates Tapia, A. N. (2016). An appraisal of the language contact profile as a tool to research local engagement in study abroad. *Study Abroad Research in Second Language Acquisition and International Education*, 1(2), 248–276. <https://doi.org/10.1075/sar.1.2.05fer>
- Forthofer, R. N., Lee, E. S., & Hernandez, M. (2007). Linear regression. In R. N. Forthofer, E. S. Lee, & M. Hernandez (Eds.), *Biostatistics* (2nd ed., pp. 349–386). Academic Press.
- Gardner, R. C. (1985). *The attitude and motivation test battery manual*. University of Western Ontario. Retrieved from <http://publish.uwo.ca/~gardner/>
- Gadsjo, J. A., Schuman, C. C., Evans, J. D., Peavy, G. M., Miller, S. W., & Heaton, R. K. (1999). Norms for letter and category fluency: Demographic corrections for age, education, and ethnicity. *Assessment*, 6(2), 147–178. <https://doi.org/10.1177/107319119900600204>
- Grendel, M. (1993). *Verlies en herstel van lexicale kennis* [Doctoral dissertation]. Radboud University Nijmegen. Retrieved January 26, 2016, from https://repository.ubn.ru.nl/bitstream/handle/2066/120134/mmubn000001_161400183.pdf
- Hansen, L. (1999). Not a total loss: The attrition of Japanese negation over three decades. In L. Hansen (Ed.), *Second language attrition in Japanese contexts* (pp. 142–153). Oxford University Press.
- Hansen, L., & Chen, Y.-L. (2001). What counts in the acquisition and attrition of numeral classifiers? *Japanese Association for Language Teaching Journal*, 23(1), 90–110. <https://doi.org/10.37546/JALTJJ23.1-5>
- Hedgcock, J. (1991). Foreign language retention and attrition: A study of regression models. *Foreign Language Annals*,

- 24(1), 43–55. <https://doi.org/10.1111/j.1944-9720.1991.tb00440.x>
- Hessel, G. (2020). Overall L2 proficiency maintenance and development among returning ERASMUS study abroad participants. *Study Abroad Research in Second Language Acquisition and International Education*, 5(1), 118–151. <https://doi.org/10.1075/sar.19011.hes>
- Huensch, A., Tracy-Ventura, N., Bridges, J., & Cuesta Medina, J. A. (2019). Variables affecting the maintenance of L2 proficiency and fluency four years post-study abroad. *Study Abroad Research in Second Language Acquisition and International Education*, 4(1), 96–125. <https://doi.org/10.1075/sar.17015.hue>
- Isurin, L., & McDonald, J. L. (2001). Retroactive interference from translation equivalents: Implications for first language forgetting. *Memory & Cognition*, 29(2), 312–319. <https://doi.org/10.3758/BF03194925>
- Jakobson, R. (1941). *Kindersprache, Aphasie und allgemeine Lautgesetze*. Almqvist & Wiksell.
- Köpke, B. (2002). Activation thresholds and non-pathological first language attrition. In F. Fabbro (Ed.), *Advances in the neurolinguistics of bilingualism: Essays in honor of Michel Paradis* (pp. 119–142). Forum Edizioni.
- Köpke, B., & Keijzer, M. (2019). Introduction to psycholinguistic and neurolinguistic approaches to language attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford Handbook of Language attrition* (pp. 63–72). Oxford University Press.
- Kuhberg, H. (1992). Longitudinal L2-attrition versus L2-acquisition, in three Turkish children-empirical findings. *Interlanguage Studies Bulletin*, 8(2), 138–154. <https://doi.org/10.1177/026765839200800203>
- Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Cybernetics and Control Theory*, 10(8), 707–710.
- Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second language acquisition. *Psychological Science*, 18(1), 29–34. <https://doi.org/10.1111/j.1467-9280.2007.01844.x>
- Linck, J. A., & Kroll, J. F. (2019). Memory retrieval and language attrition: Language loss or manifestation of a dynamic system? In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 88–97). Oxford University Press.
- Luo, L., Luk, G., & Bialystok, E. (2010). Effect of language proficiency and executive control on verbal fluency performance in bilinguals. *Cognition*, 114(1), 29–41. <https://doi.org/10.1016/j.cognition.2009.08.014>
- Mehotcheva, T. H. (2010). *After the fiesta is over: Foreign language attrition of Spanish in Dutch and German Erasmus students* [Doctoral dissertation], Pompeu Fabra. Retrieved May 16, 2018, from <https://www.tdx.cat/bitstream/handle/10803/37468/tm.pdf>
- Mehotcheva, T. H., & Köpke, B. (2019). Introduction to L2 attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 331–348). Oxford University Press.
- Mehotcheva, T. H., & Mytara, K. (2019). Exploring the impact of extralinguistic factors on L2/FL attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 349–363). Oxford University Press.
- Mickan, A., McQueen, J. M., & Lemhöfer, K. (2019). Bridging the gap between second language acquisition research and memory science: The case of foreign language attrition. *Frontiers in Human Neuroscience*, 13, 397. <https://doi.org/10.3389/fnhum.2019.00397>
- Mickan, A., McQueen, J. M., & Lemhöfer, K. (2020). Between-language competition as a driving force in foreign language attrition. *Cognition*, 198, 104218. <https://doi.org/10.1016/j.cognition.2020.104218>
- Murtagh, L. (2003). *Retention and attrition of Irish as a second language* [Doctoral dissertation], University of Groningen. Retrieved May 16, 2018, from <https://www.rug.nl/research/portal/files/2999446/thesis.pdf>
- Nikitina, L., & Furuoka, F. (2005). Integrative motivation in a foreign language classroom: A study on the nature of motivation of the Russian language learners in universiti Malaysia Sabah. *Jurnal Kinabalu, Jurnal Perniagaan & Sains Sosial*, 11, 23–34.
- Olshain, E. (1989). Is second language attrition the reversal of second language acquisition? *Studies in Second Language Acquisition*, 11(2), 151–165. <https://doi.org/10.1017/S027263100000589>
- Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. John Benjamins. <https://doi.org/10.1075/sibil.18>
- R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- Rosselli, M., Ardila, A., Salvatierra, J., Marquez, M., Matos, L., & Weekes, V. A. (2002). A cross-linguistic comparison of verbal fluency tests. *International Journal of Neuroscience*, 112(6), 759–776. <https://doi.org/10.1080/00207450290025752>
- Schmid, M. S. (2007). The role of L1 use for L1 attrition. In B. Köpke, M. S. Schmid, M. Keijzer, & S. Dostert (Eds.), *Language attrition: Theoretical perspectives* (pp. 135–153). John Benjamins.
- Schmid, M. S. (2019). The impact of frequency of use and length of residence on L1 attrition. In M. S. Schmid & B. Köpke (Eds.), *The Oxford handbook of language attrition* (pp. 288–303). Oxford University Press.
- Schmid, M. S., & Mehotcheva, T. (2012). Foreign language attrition. *Dutch Journal of Applied Linguistics*, 1(1), 102–124. <http://doi.org/10.1075/dujal>
- Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, 5, 772. <https://doi.org/10.3389/fpsyg.2014.00772>
- Tomiyama, M. (2008). Age and proficiency in L2 attrition: Data from two siblings. *Applied Linguistics*, 30(2), 253–275. <https://doi.org/10.1093/applin/amn038>
- Tulloch, B., & Ortega, L. (2017). Fluency and multilingualism in study abroad: Lessons from a scoping review. *System*, 71, 7–21. <http://doi.org/10.1016/j.system.2017.09.019>
- Wang, X. (2010). *Patterns and causes of attrition of English as a foreign language* [Doctoral dissertation], Shandong University. Retrieved June 3, 2020, from <https://www.let.ru.nl/languageattrition/Papers/Wang2010.pdf>
- Welters, B. (1988). *The attrition of French as a foreign language* [Doctoral Dissertation], Katholieke Universiteit Nijmegen. Retrieved May 16, 2018, from https://repository.ubn.ru.nl/bitstream/handle/2066/113589/mmubn000001_071056505.pdf

- Williams, S., & Hammarberg, B. (1998). Language switches in L3 production: Implications for a polyglot speaking model. *Applied Linguistics*, 19(3), 295–333. <https://doi.org/10.1093/applin/19.3.295>
- Xu, X. (2010). *English language attrition and retention in Chinese and Dutch university students* [Unpublished doctoral dissertation], University of Groningen.
- Yang, J.-S. (2016). The effectiveness of study-abroad on second language learning: A meta-analysis. *The Canadian Modern Language Review*, 72(1), 66–94. <http://doi.org/10.3138/cmlr.2344>
- Yoshitomi, A. (1999). On the loss of English as a L2 by Japanese returnee children. In L. Hansen (Ed.), *Second language attrition in Japanese contexts* (pp. 80–111). Oxford University Press.