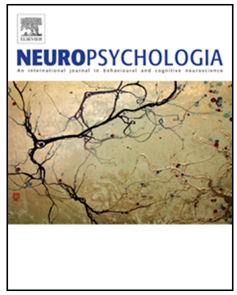
Journal Pre-proof



Electrophysiological evidence for cross-language interference in foreign-language attrition

Anne Mickan, James M. McQueen, Beatrice Valentini, Vitoria Piai, Kristin Lemhöfer

PII: S0028-3932(21)00046-4

DOI: https://doi.org/10.1016/j.neuropsychologia.2021.107795

Reference: NSY 107795

To appear in: *Neuropsychologia*

Received Date: 19 August 2020

Revised Date: 18 December 2020

Accepted Date: 7 February 2021

Please cite this article as: Mickan, A., McQueen, J.M., Valentini, B., Piai, V., Lemhöfer, K., Electrophysiological evidence for cross-language interference in foreign-language attrition, *Neuropsychologia*, https://doi.org/10.1016/j.neuropsychologia.2021.107795.

Credit author statement

Anne Mickan: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing - Original Draft, Visualization, Project administration, Funding acquisition

Beatrice Valentini: Software, Validation, Formal analysis, Investigation, Data Curation, Writing - Review & Editing

Vitória Piai: Formal analysis, Writing - Review & Editing, Supervision

James M McQueen.: Conceptualization, Methodology, Writing - Review & Editing, Supervision

Kristin Lemhöfer: Conceptualization, Methodology, Writing - Review & Editing, Supervision

Electrophysiological evidence for cross-language interference in foreign-language attrition

Anne Mickan[a,b,c], James M. McQueen[a,b], Beatrice Valentini[d], Vitoria Piai[b,e], & Kristin Lemhöfer[b]

[a] Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

[b] Radboud University, Donders Institute for Brain, Cognition and Behaviour, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands

[c] International Max Planck Research School for Language Sciences, Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands

[d] University of Geneva, Faculty of Psychology and Social Sciences, 40 Boulevard Pont d'Arve, 1205 Geneva, Switzerland

[e] Radboudumc, Donders Centre for Medical Neuroscience, Department of Medical Psychology, Nijmegen, The Netherlands

Corresponding author:

Anne Mickan (a.mickan@donders.ru.nl), Montessorilaan 3, 6525 HR Nijmegen, The Netherlands

Declarations of interest: none

Journal Pre-proof

Abstract

Foreign language attrition (FLA) appears to be driven by interference from other, more recently-used languages (Mickan et al., 2020). Here we tracked these interference dynamics electrophysiologically to further our understanding of the underlying processes. Twenty-seven Dutch native speakers learned 70 new Italian words over two days. On a third day, EEG was recorded as they performed naming tasks on half of these words in English and, finally, as their memory for all the Italian words was tested in a picture-naming task. Replicating Mickan et al., recall was slower and tended to be less complete for Italian words that were interfered with (i.e., named in English) than for words that were not. These behavioral interference effects were accompanied by an enhanced frontal N2 and a decreased late positivity (LPC) for interfered compared to not-interfered items. Moreover, interfered items elicited more theta power. We also found an increased N2 during the interference phase for items that participants were later slower to retrieve in Italian. We interpret the N2 and theta effects as markers of interference, in line with the idea that Italian retrieval at final test is hampered by competition from recently practiced English translations. The LPC, in turn, reflects the consequences of interference: the reduced accessibility of interfered Italian labels. Finally, that retrieval ease at final test was related to the degree of interference during previous English retrieval shows that FLA is already set in motion during the interference phase, and hence can be the direct consequence of using other languages.

Keywords: foreign language attrition, theta oscillations, N2, inhibition, competition

Electrophysiological Evidence for Cross-Language Interference in Foreign-Language Attrition

Most people who have learned a foreign language will be familiar with the frustrating feeling of losing access to that language over time, no matter how much effort they put into learning it in the first place. Why this happens, and why the so-called attrition process appears to be so inevitable, is a long-standing issue in the language sciences. Recent research suggests that foreign language attrition can be the direct consequence of using and speaking other languages (e.g., Levy et al., 2007; Mickan et al., 2020). Mickan et al. (2020), for example, showed that the mere act of retrieving words in either a native or a foreign language hampers subsequent access to translation equivalents in another foreign language, and that these interference effects are long-lasting. The neural correlates of these processes and, hence, their exact contribution to foreign language attrition, however, are still unknown. The current study aims at filling this gap. Building on Mickan et al. (2020), we seek to establish the electrophysiological correlates of between-language interference. The electroencephalogram (EEG) provides a different way of looking at the attrition process, and, crucially, allows us to understand precisely when in time these interference effects emerge. In doing so, we hope to shed light on the temporal dynamics of the underlying mechanisms of interference-based foreign language (FL) forgetting.

1.1 Between-Language Competition and Inhibition as Driving Forces in FL Attrition

To investigate why we forget foreign language vocabulary, Mickan et al. (2020) took inspiration from the domain-general memory literature. According to research in this domain, forgetting is not just a by-product of disuse and the passage of time, but instead often (if not always) the consequence of competition and thus interference between related memories (e.g., Anderson et al., 1994; Müller & Pilzecker, 1900; Underwood, 1957). A classic example of forgetting through competition is the so-called retrieval-induced forgetting (RIF) phenomenon (Anderson et al., 1994). When participants selectively practice a subset of previously learned category-exemplar pairs (e.g., FRUIT-banana), this retrieval practice has been shown to impair later access to other exemplars from the practiced categories (i.e., other exemplars from the category FRUIT), but not to exemplars from unpracticed categories (e.g., exemplars from the category FURNITURE). In other words, retrieving information can make us forget related, competing information.

In two experiments, Mickan et al. (2020) asked whether RIF, and thus forgetting through competition and interference, is applicable to the FL attrition context. Do we forget words from foreign languages because we use other languages instead? Reason to believe that this might be the case comes from studies on bilingual word production that show that translation equivalents in multiple languages tend to be simultaneously activated (Kroll et al., 2008). While this co-activation can be beneficial (e.g., Costa et al., 2000), it can also result in competition between languages and hence in interference (e.g., Colomé, 2001; Hermans et al., 1998). Just as banana and apple compete with one another by virtue of being connected to the same overarching semantic category, so can translation equivalents interfere with one another when cued with their shared concept. Inspired by these parallels, Mickan and colleagues asked whether the between-language competition that is sometimes observed during short-term, online processing also has long-term ramifications and hence whether it makes for a plausible mechanism behind FL attrition.

Participants in Mickan et al.'s (2020) experiments first learned a set of new L3 Spanish words. One day later, they were asked to repeatedly retrieve half of those in either L1 Dutch or L2 English. Finally, after a delay of 20 minutes, participants were tested again on all originally learned Spanish words. Naming performance in this final test showed that people were significantly worse at recalling words that they had just named in English or Dutch:

they made more mistakes and were slower to recall interfered compared to not interfered items in Spanish. Lexical retrieval of translation equivalents in a different language can thus make you forget words from a (recently learned) foreign language. In reaction times, this effect persisted until a week after interference induction, thus providing the first evidence for true long-term effects of between-language interference, and hence establishing it as a plausible mechanism of foreign language attrition.

These and comparable retrieval-induced forgetting effects in the memory literature tend to be explained through inhibition processes (Anderson et al., 1994). In the language case, specifically, Mickan et al. (2020) reasoned that during the retrieval of English words in the interference phase of their experiment, the recently learned Spanish words competed for access with their English translation equivalents and hence that they had to be inhibited for successful retrieval of the latter. Assuming that this inhibition is long-lasting, it should result in a competition disadvantage for the suppressed Spanish items at delayed final test (i.e., after 20 minutes). In order to retrieve the Spanish labels, their inhibition first needs to be lifted and competition from their recently retrieved English competitors needs to be overcome, which takes time and hence slows down or entirely blocks retrieval. Words in the no-interference condition, which were not retrieved in English and hence did not need to be inhibited in Spanish, should consequently be easier to retrieve and experience less interference from English competitors at final test than items that were interfered with, which explains why the former were recalled faster and more accurately than the latter. Between-language competition effects had previously mostly been established between L1 and L2 and short-term, that is, in experimental designs that document these effects within individual (or pairs of) trials (i.e., immediately rather than after a delay; though see Branzi et al., 2014; Misra et al., 2012). The results from Mickan and colleagues suggest that between-language competition also unfolds between two foreign languages (as well as between L1 and L3) and that it can have long-term ramifications.

In the present paper, we seek to replicate the main effects of interference reported in Mickan et al. (2020), but crucially aim to further our understanding of the underlying cognitive mechanisms driving this behavioral effect. To that end, we measured EEG activity both during the interference phase and during the final test phase and asked whether we could track the competition and inhibition dynamics that are often called upon in explaining the behavioral between-language interference effects. Instead of testing for interference on L3 Spanish, we used Italian as L3 in the current study, and the interference phase consisted only of L2 English retrieval practice (leaving out the L1 Dutch group from Mickan et al.). Dutch native speakers thus first learned a set of words in L3 Italian, and subsequently, a day later, retrieved half of them in L2 English, before being tested again on all originally learned Italian words.

We expected to observe neural correlates of interference and inhibition both at final test and during the interference phase. Moreover, in looking not only at the outcome of such interference (i.e., the final test), but also at the moment in time when forgetting is supposedly induced (i.e., the interference phase), we aimed at testing the assumption that activity during the earlier phase is directly related to performance at final test. If competition and inhibition during the interference phase indeed predict retrieval ability at final test, we should be able to observe more competition/inhibition for items that are later on slower to retrieve (i.e., harder to recall at final test) compared to items that are fast to retrieve at final test.

1.2 Stimulus-Locked Neural Markers of Interference, Competition and Inhibition

1.2.1 Evidence from Event-Related Potentials

In event-related potentials (ERPs) in the EEG, inhibition and interference are commonly associated with an early anterior negative deflection, the so-called N2 component.

Maximal over frontal electrode sites and peaking between 200 and 350 ms (time-locked to stimulus presentation), this component has frequently been observed in studies using the language switching paradigm, where people alternate between naming pictures in their L1 and L2 prompted by a language cue. In those situations, it is common to find an enhanced N2 for switch trials, where the language of naming differs from the previous trial, compared to repeat trials where the language remains the same (Jackson et al., 2001; Zheng et al., 2020; but see Christoffels et al., 2007, for a larger N2 for repeat compared to switch trials). These N2 switch costs are typically interpreted to reflect inhibition of the non-target language, in line with interpretations of comparable N2 findings in non-linguistic tasks that require inhibition of a prepotent response (e.g., the Go-NoGo-task; see Folstein & Van Petten, 2008, for a review). Some researchers have instead argued that the N2 is a signature of response conflict, rather than evidence for the resolution of that conflict (i.e., through inhibition of interfering responses or boosting of target responses; Nieuwenhuis et al., 2003). Crucially though, both interpretations assume that it is an indicator for the presence of interference, and hence is a viable candidate for a neural correlate of interference-based foreign language attrition.

It should be noted that most evidence for language switch N2 effects comes from mixed-language switching paradigms, which differ in design from the current study in important ways. First of all, traditional language switching studies test for inhibition on a global, whole-language level rather than locally on the item level: they ask what naming a picture in, for example, L1 does to subsequent naming of any other picture in L2, rather than to naming of its L2 translation equivalent. Moreover, they observe the effects of language switching from one trial to the next, but not their potential long-term effects (though see Branzi et al., 2014; Misra et al., 2012; Wodniecka et al., 2020; reviewed in detail in the discussion section). It remains to be seen whether the sustained, local interference / inhibition effects underlying foreign language attrition are reflected in the same N2 modulation as the short-lived, global effects in mixed-language switching studies. Finally, language switching studies differ from our study in that they target switching between two already known languages, namely L1 and L2, but not, at least to our knowledge, switching between two foreign languages, of which one has just recently been learned. The neural correlates of the consequences of naming in one foreign language on subsequent, delayed naming in another, just recently learned foreign language, as studied here, thus remain to be investigated. If the effects observed by Mickan et al. (2020) are caused by language interference and inhibition and assuming that the N2 reflects these processes not just globally, but also locally, we should expect modulations of the N2 component in the EEG during both the interference and the final test phase.

Another component that is sometimes reported in language switching studies is the LPC, a late positive component with a posterior parietal topography, occurring between 300 and 900 ms post stimulus onset. Just like the N2, the LPC is bigger for switch compared to non-switch trials and has hence been interpreted as a continuation thereof, indexing the after-effects of language interference and inhibition (Jackson et al., 2001). This component is not always found, and in fact not even always inspected (the time window for analysis in switching studies is typically limited to the first 500 ms post stimulus presentation), and hence it is unclear how robust this signature is. Importantly, this ‘switching LPC’ is not to be confused with the much more frequently reported LPC in the memory literature. The ‘memory LPC’ is thought to reflect long-term episodic recognition processes. During retrieval, it has been reliably found to be bigger for old compared to new items (for a review, see Rugg & Curran, 2007). In contrast to the switching LPC, which appears to be enhanced during retrieval of items on which there is more interference (i.e., from a non-target language, namely on switch trials), the memory LPC is found to be stronger for items where retrieval is more accurate and successful (e.g., Finnigan et al., 2002; Rugg et al., 1995; Wilding, 2000), and hence would be expected to show the opposite pattern, that is to be enhanced in trials where interference is low rather than high. Note, however, that this memory LPC is typically reported in recognition paradigms, rather than during productive recall. It remains to be seen whether our interference manipulation influences either of these late positive effects and, hence, whether the LPC is also a marker of foreign language attrition or not.

1.2.2 Evidence from Neuronal Oscillations

In the frequency domain, interference has been consistently associated with power increases in the theta band (4-7Hz) of the EEG signal. Evidence comes, for example, from studies using tasks with response conflicts, such as the Go-NoGo or Stroop tasks, where theta power (time-locked to stimulus onset) is enhanced in the conflicting compared to the not (or less) conflicting condition (Hanslmayr et al., 2008; Nigbur et al., 2011). These theta effects occur anywhere within the first 1000 ms post stimulus presentation, tend to have a mid-frontal scalp distribution and are thought to reflect interference from alternative responses, and possibly the recruitment of executive control processes to overcome this interference.[1]

In the language domain, theta power has been linked to semantic interference: naming a picture with a semantically related, same-language distractor displayed on top triggered more theta activity than naming a picture with a semantically unrelated, and hence less interfering distractor on top (Piai et al., 2014). Between-language interference, time-locked to the presentation of a stimulus, as targeted in this paper, however, has not yet been linked to

[1] Note that these theta effects are different from the theta effects that have been linked to successful memory encoding; these will not be discussed here any further (e.g., Klimesch et al., 1996).

theta power increases. To our knowledge, there are no studies on the oscillatory dynamics of stimulus-induced between-language competition in bilingual word production.

Further evidence for theta as a marker for interference magnitude comes from memory studies on retrieval-induced forgetting (RIF; e.g., Ferreira et al., 2014; Hanslmayr et al., 2010; Staudigl et al., 2010). These studies typically contrast competitive and non-competitive interference conditions during the retrieval of previously learned category-exemplar associations. Staudigl et al. (2010), for example, had participants either actively retrieve a subset of previously studied exemplars from a given category, or passively restudy category-exemplar pairs. In the active retrieval condition, the presentation of the category cue activates other exemplars which compete with selection of the to-be-retrieved exemplar, while no such competition and interference emerges when passively viewing category-exemplar pairs. In line with the idea that theta is a marker for interference magnitude, theta power was found to be increased during retrieval in the active retrieval task as compared to the passive exposure task. Changes in theta power from the first to the second round of active competitive retrieval were furthermore found to be related to later forgetting. Forgetting in RIF studies is measured in a final test on all originally learned category-exemplar pairs, both those intermittently retrieved or restudied and those not part of the interference phase at all. Behaviorally, Staudigl et al. (2010) only observed forgetting for exemplars whose competitors (i.e., other exemplars from the same category) were actively retrieved in the interference phase, but not for exemplars whose competitors were only restudied. Crucially, the magnitude of forgetting was positively correlated with the decrease in theta from the first to the second round of retrieval practice, suggesting that interfering competitors were suppressed during competitive retrieval and that the amount of this suppression was related to later forgetting.

Next to oscillations, EEG RIF studies also sometimes report ERPs (Ferreira et al., 2014; Hanslmayr et al., 2010; Johansson et al., 2007). Unlike the theta effects, the ERP signatures they report vary considerably from study to study though, ranging from prolonged positivities (Johansson et al., 2007) to a combination of short-lived posterior negativities and anterior positivities (Hanslmayr et al., 2010; Ferreira et al., 2014) for competitive compared to non-competitive retrieval. It should be noted though that comparisons in EEG RIF studies are often between entirely different tasks (e.g., active retrieval vs. passive restudy), making it unclear to what extent their theta and ERP signatures reflect only competition or also other task-related differences between conditions. Even when the comparison is between two active retrieval tasks though, as in Hanslmayr et al. (2010), their stimuli (category-exemplar pairs) and task design (covert rather than overt retrieval) make the comparison to the present study difficult. Given these design differences and the inconsistent ERP signatures RIF studies report, it is questionable how relevant they are for hypothesis formulation for the present study. For ERPs, we consider the N2 component to be much more likely given its reliable presence in studies that require switching between languages in overt picture naming paradigms. For theta oscillations, there is no evidence for their involvement in competitive bilingual lexical retrieval to this point, and hence it will be interesting to see whether they are implicated in the type of between-language competition and interference that is supposedly at play in foreign language attrition, or not.

1.3 The Present Study

To sum up, the present study investigates the neural correlates of foreign language attrition. Building on previous behavioral studies, we seek converging neural evidence for between-language interference and inhibition as driving forces behind foreign language vocabulary forgetting. To that end, Dutch native speakers first learned 70 new Italian words over the course of two consecutive days. On a third and last day, they were asked to retrieve half of the learned words in English, a foreign language they already knew, and were subsequently tested on all originally learned Italian words. We chose English as interference language because Mickan et al. (2020) had found that foreign languages tend to be stronger interferers than the L1. We measured EEG during all sessions on the third day, that is both during the picture naming tasks in the interference phase and at final test.

Behaviorally, we expected to replicate Mickan et al. (2020) despite the change in language (Italian rather than Spanish), the extended memory set (70 instead of 40 to be learned words) and the fact that the learning session was spread over two rather than just one day. We thus predicted to observe more errors and slower naming responses to interfered than not interfered words at final test in Italian. Critically, though, based on the EEG literature reviewed in the previous sections, we also expected those behavioral effects to be accompanied by possibly more theta power and most likely an increased N2 component for interfered items at final test. In line with how these signatures are typically interpreted, we hypothesize that theta indexes the interference that the Italian items experience from the recent practice of their English translation equivalents and that the N2 reflects the higher need for inhibition of the latter to resolve this interference. We had no clear expectations with regard to the LPC.

We were also interested in the interference phase itself, when forgetting is supposedly induced. Here our comparison of interest concerns only the items in the interference condition (as the other items did not occur in this phase). If cognitive control dynamics during the interference phase are responsible for performance deficits at final test, we should observe more evidence for such processes on items that are later more difficult to recall at final test. To that end, we analyzed, per participant, their trials in the interference phase based on a median split of their naming latencies at final test. We expected that words that took participants longer to recall at final test would show an enhanced N2 and stronger oscillations in the theta range during the English interference phase than words that participants were faster to retrieve at final test in Italian.

2. Methods

2.1 Participants

Thirty Dutch native speakers were recruited via the Radboud University participant pool. One failed to reach the learning criterion on day 2 (see section 2.3 for details), and hence had to be excluded from the remainder of the study. Two additional participants had to be excluded from analysis because they had too many EEG artifacts (see section 3.1 for details). The remaining 27 participants (18 female) were between 18 and 26 years old (*M* = 21.07; *SD* = 2.00). All of them were right-handed, had normal or corrected-to-normal vision, and reported no history of language-related or neurological impairments. For the analysis of the interference phase EEG recordings, one of these 27 participants had to be discarded because of technical failure (and hence missing data) during this part of the experiment.

Before coming to the lab, participants were asked to fill in an online language background questionnaire. This was done to ensure that our participants had no (or minimal) prior knowledge of Italian. Only one participant reported prior knowledge of Italian. He had only just started learning Italian on Duolingo a month prior to participating in the study, and judged his Italian as very poor (1 out of 7). He was deemed sufficiently inexperienced with Italian to still be included in the experiment.

As also established through this online questionnaire, Dutch was our participants' only mother tongue and English was the first learned foreign language for all participants. Table 1 summarizes our participants' frequency of use and proficiency self-ratings in English, as well as their performance on the English LexTALE, a standardized lexical-decision based vocabulary test (Lemhöfer & Broersma, 2012). Other languages participants spoke included most prominently German, French and Latin. We refer to Italian as an L3 because it was learned after L2 English. For some participants, Italian was actually L4 or even L5, but we stick to L3 for simplicity.

Participants gave informed consent and received either course credit or vouchers for their participation (10€/h). The study was approved by the Ethics Committee of the Faculty of Social Sciences, Radboud University.

Table 1

Participant characteristics

	M	*SD*	*range*
English AoA	10.41	1.19	8-12
English LoE (years)	9.67	3.09	5-16
Frequency of Use (min/day)			
-Speaking	29	62	0-300
-Listening	160	132	0-480
-Reading	87	74	0-270
-Writing	43	67	0-240
Proficiency[a]			
-Speaking	5.52	1.09	4-7
-Listening	5.93	1.00	4-7
-Reading	5.97	0.90	4-7
-Writing	5.37	1.25	3-7
English LexTALE (in %)	74	14	43-95

Note. *M* = mean; *SD* = standard deviation; AoA = Age of acquisition; LoE = length of exposure (i.e., amount of years participants had been learning English). [a]Proficiency self-ratings were given on a scale from 1 (very bad) to 7 (like a native speaker)

2.2 Materials

Participants learned 70 Italian nouns referring to concrete, everyday objects or animals (see Appendix A for the list of words). All words were non-cognates between Italian, Dutch and English, and were between 2 and 4 syllables long in Italian (M = 2.69, SD = 0.50) and between 1 and 3 syllables long in English (M = 1.33, SD = 0.67). Their corresponding

Dutch lemma frequencies ranged from 0 to 180 per million (*M* = 25.11, *SD* = 37.30; CELEX, Baayen et al., 1995). Pictures for each of the words were chosen from Google (www.google.com) and the BOSS database (Brodeur et al., 2010). They were photographs of the respective object or animal centered on a white background (6x6 cm) and adjusted for size so that they occupied a maximum of 400px in either width or length. Finally, each noun was recorded by a female Italian native speaker from Rome (Italy).

These 70 words were subdivided into two subsets of thirty-five words each: one of those two subsets was later interfered with, that is retrieved in English on day 3, while the other was not (see Appendix A for each word's set assignment). Which set received interference was counterbalanced across participants. Importantly, though, the two subsets were matched in terms of word length in both Italian and English, Dutch frequency, within-set phonological similarity as assessed via Levenshtein distances (Levenshtein, 1966), and within-set semantic similarity (expressed as a distance value derived from semantic vectors with smaller values corresponding to high semantically similarity, as described in Mandera et al., 2017) (see Table 2 for averages of these values per set).

For the interference phase, 35 filler items to be named in English were chosen in addition to the 35 experimental items that would receive interference. Filler items were not analyzed, and were merely included to disguise the fact that only half of the originally learned experimental items were part of the interference session. Filler items were nevertheless matched to the experimental items in terms of English word length (*M* = 1.43, *SD* = 0.50, range = 1-2) and Dutch frequency (*M* = 1.20, *SD* = 0.55, range = 0-2.24).

Table 2

Item characteristics

	Set 1			Set 2		
	M	*SD*	*range*	*M*	*SD*	*range*
Italian word length (in syllables)	2.66	0.68	2-4	2.72	0.66	2-4
English word length (in syllables)	1.26	0.44	1-2	1.40	0.55	1-3
Dutch Celex log frequency	0.97	0.66	0-2.14	1.09	0.54	0-2.26
Dutch Celex per million frequency	24.63	34.36	0-137	25.60	40.53	0-180
Semantic distance[a]	0.81	0.17	0-1.09	0.81	0.17	0-1.05
Italian Levenshtein distance	6.42	1.64	2-11	6.28	1.51	1-10

Note. M = mean, *SD* = standard deviation. Which set received interference was counterbalanced across participants. [a]Semantic similarity was assessed via semantic vectors, as described in Mandera et al. (2017). Small values reflect higher semantic overlap.

2.3 Procedure

The study consisted of three consecutive testing days (see Figure 1 for a schematic overview). On the first two of those, participants were asked to learn 70 Italian words via a mix of receptive and productive tasks with feedback. Learning success was established at the end of each day via a posttest without feedback (see below for details). No EEG was acquired during either of those two days. The third and last day started with the so-called interference phase, during which participants were asked to retrieve half of the learned words in English, and ended with a final test in Italian on all originally learned words. To avoid confusion, we refer to the Italian recall test on day 3 as ‘final test’, rather than posttest; the word ‘posttest’ is used to refer to the Italian recall tests at the end of day1 and 2. EEG was acquired during the entire session on day 3, that is both during the interference phase and the final test in Italian. Below, we will describe the tasks of the various phases of the experiment in more detail.

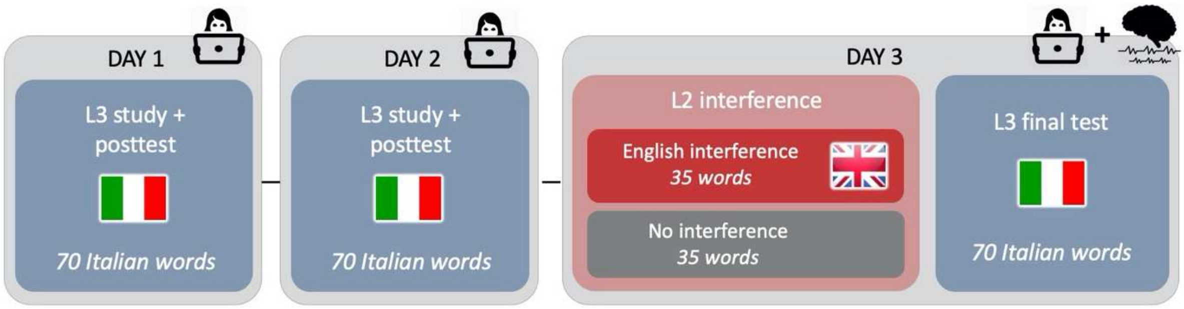


Figure 1. Schematic overview over experimental set-up.

All tasks were administered using Presentation (Version 19.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com) on a Dell T3610 computer (3,7Ghz Intel Quad Core, 8GB RAM, Windows 7, monitor: BenQ XL 2420Z, 24-in, 1920x1080 pixels, 120 Hz refresh rate). All audio stimuli were presented to the participants via headphones (Sennheiser HD201), and all oral responses were recorded via a microphone (Shure 16a) in WAV format using a Behringer X-Air XR18 digital mixer.

On all days, participants were tested individually in a quiet room. For the behavioral sessions on days 1 and 2, the experimenter sat in a room next to the participant's room. The door between these two rooms was kept open at all times for efficient communication, and for the experimenter to be able to code the participant's responses. On day 3, for the EEG session, the experimenter also sat in an adjacent room, this time the door was kept shut during the experiment and communication between experimenter and participant was done via microphone.

2.3.1 Day 1 – Italian Learning Phase 1

The learning phase consisted of a series of receptive and productive tasks that started out easy, and got progressively more engaging and difficult. The learning phase on day 1 started with a familiarization round, during which participants listened to and saw (written versions of) each of the 70 words on screen, as well as their corresponding pictures once.

Participants clicked through the pictures at their own pace. Next to acquainting themselves with the items, they were also instructed to let the experimenter know if they already knew any of the Italian words. Italian words that were already known to a participant were later excluded from analysis (also see section 3.1; the average number of excluded items across participants was $M = 0.67$, $SD = 1.71$, range = 0-8). This initial familiarization round was followed by two rounds of a two-alternative forced choice task, in which participants saw all 70 pictures twice, each time with two Italian labels from the list of to-be-learned words underneath. Participants were asked to choose the word that matched the picture they saw by clicking on it with a mouse. They then received automatic feedback on their performance (a green or red square around the picture for correct or incorrect responses respectively, accompanied by the correct word underneath the picture and its corresponding audio). After the feedback, the next trial started automatically. In the second round of this task, before seeing the two labels, participants were asked to guess the Italian word for the picture and say it out loud. This was done to start engaging them more actively with the words. The experimenter initiated the appearance of the two labels after a participant had made an attempt at naming the picture, and the rest of the trial continued as in the first round.

Next, participants completed two rounds of a word completion task. They saw each of the pictures together with their respective first syllables (or first grapheme for monosyllabic words) and were asked to complete the word out loud. The experimenter coded their answers for correctness (either as fully correct, fully incorrect or partially correct), which initiated feedback (identical to the feedback in the multiple-choice task, a green frame was displayed only for fully correct answers). From this task onwards, participants could decide for themselves when to continue with the next trial; they were thus allowed as much time as they needed to process the feedback. The word completion task was followed by a writing task: participants saw each picture once and were asked to write down (on a piece of paper) the

Italian word for the picture. They then hit the enter key, which initiated the visual presentation of the correct Italian label and its spoken form. Based on this feedback, participants then corrected themselves, when necessary, by writing down the correct word on the same piece of paper. They were instructed to write each word on a new piece of paper, and to turn over each piece after use so that they would not be able to see their earlier responses.

The first day ended with two rounds of picture naming. The first of those rounds was still with feedback: participants saw each picture once, had to name it and received feedback initiated by the experimenter. Words were again coded as fully correct, fully incorrect or partially incorrect, and feedback was also again a green (for fully correct answers) or red screen (for partially and fully incorrect answers) together with the correct label (presented visually and auditorily). The second round served as a posttest, to establish which words had already been learned. During this last round, participants no longer received feedback.

2.3.2 Day 2 – Italian Learning Phase 2

Day 2 (mean hours between day 1 and day 2 = 23.17, *SD* = 2.78, range =18-29) of the learning phase started with another round of the word completion task. The set-up was identical to the word completion task on day 1. The remainder of the session was spent with picture naming tasks, similar in set-up to the picture naming tasks on day 1. Participants named each picture at least twice with feedback. If a participant knew all words in those two rounds, they proceeded to one more round of naming with feedback, followed by one final round of naming without feedback (i.e., the second Italian posttest, identical in set-up to the posttest on day 1). If a participant did not know all words during the first two rounds of picture naming, the unknown words were repeated until he/she had named all pictures correctly in at least two consecutive rounds. Each repetition round consisted of at least ten pictures: if a participant only had two pictures left to learn, they would thus get both these pictures, but also eight already known pictures to name. This was done to ensure sufficient difficulty in naming even when there were only few items left to be learned. Repetitions of already known words were counterbalanced, such that each word was repeated approximately equally many times.

Throughout the entire learning session, taking day 1 and 2 together, participants saw each picture minimally 14 times (*M* = 15.34, *SD* = 1.03, abs. range = 14-30). Both sessions took a maximum of 1.5 hours. If on day 2 participants were still on the adaptive picture naming task after 1 hour and 15 minutes, this task was stopped by the experimenter, and the remaining two rounds of picture naming were administered. Participants were required to learn a minimum of 50 out of the 70 words (spread equally over conditions) to be able to continue to day 3. As mentioned in the Participants section, all but one participant reached this criterion.

The order of items in all learning tasks was participant- and task-specific: to avoid order effects during learning, pictures were presented in a different, random order in each task. To keep the distance between item repetitions constant within a task though, the order of items was identical for consecutive rounds of a single task. Due to the set-up of the tasks, there were never more than two identical rounds in a row. For the two posttests (one on each day), six lists were created, making sure that no more than three items from the same condition (interfered / not interfered) followed in immediate succession, and that half of the participants started each posttest with an item from the interference condition and the other half with an item from the no interference condition. The items in lists 4 to 6 followed the reversed order of the items in lists 1 to 3. Each participant got a different list for each of the two posttests, but never two reversed lists (e.g., never 1 and 4, or 2 and 5). All tasks in the learning phase were also used in Mickan et al. (2020). See that paper for further procedural details.

2.3.3 Day 3 – English Interference Phase and Final Italian Test

2.3.3.1 Interference Phase. Day 3 (mean hours between day 2 and day 3 = 24.46; *SD* = 3.62, range = 17-32) started with an interference phase, during which participants saw the pictures corresponding to half of the learned Italian words, as well as 35 filler pictures, and had to retrieve the names of the pictures in English. In total, they saw each picture nine times: Once during an initial (English) familiarization task with feedback, four times during a picture naming task without feedback and another four times during a letter search task without feedback. EEG data were acquired during all these tasks but only those from the picture naming tasks were analyzed. Furthermore, out of the picture naming tasks, only the first and last rounds were analyzed (see Results section for details).

In the familiarization task, each trial started with a fixation cross presented on the screen for 1500 ms, followed by a blank screen for 500 ms, followed in turn by the picture together with the first syllable (or the first grapheme in case of monosyllabic words) of its corresponding English label. We chose to present syllables rather than the initial letter to make naming easier. The picture and text were displayed for 2000 ms. Participants were instructed to withhold their response during this delay period. In the subsequent picture naming tasks, this delay would serve as the time window for EEG analysis and hence needed to be as free of movement artifacts as possible. Given that this delay is the same for all words regardless of which condition they belong to, differences between conditions should be unaffected by it. Data of the familiarization task were not analyzed, and hence the delay was not strictly necessary here, but we included it anyway to familiarize participants with the task timing. After these 2000 ms, a question mark appeared on the screen prompting the participant to give their response, that is, name the picture in English. The experimenter coded their answers for correctness (fully correct, fully incorrect or partially correct as during the learning phase), and in doing so initiated a feedback screen, which unlike the feedback in the learning tasks only contained the intended, correct English label for the picture, but no green or red screen and also no audio. If a participant had been unable to name a picture in English, the experimenter asked whether they at least recognized the word on screen, or whether it was indeed an entirely unknown English word for the participant. If a participant indicated recognizing the picture, the item was subsequently marked as known rather than unknown. Only truly unknown words, that is target words that were neither named correctly nor recognized by a participant, were later excluded from analysis in all tasks (see section 3.1; average number of excluded items: M = 1.33, SD = 2.10, range = 0-9). The feedback screen remained visible until the experimenter confirmed or changed the correctness coding. The next trial then started automatically.

The picture naming task also started with a 1500 ms fixation cross, followed by a 500 ms blank screen, and finally the picture for 2000 ms. Participants were again instructed to withhold their response during this delay window, and to blink as little as possible. The experimenter again coded responses for correctness, but participants did not receive feedback, and the experimenter's button press immediately initiated the next trial.

In the letter search task that followed, participants had to decide whether or not the English word for the picture contained a certain letter. For each round, participants got a new letter (one of R, L, T, or N). A trial started with a 500 ms fixation cross, followed by a 250 ms blank screen, and finally the picture, which remained on screen until a participant pressed a button (right button for yes, left button for no), or for a maximum of 10 seconds. Participants did not receive feedback on their performance.

In order to make the interference phase less monotonous, we split the picture naming and letter search tasks, such that participants first underwent two rounds of picture naming, followed by two rounds of letter searching (letters R and L), followed by two more rounds of picture naming and two more rounds of letter searching (letters T and N). The presentation order of items in the interference tasks was semi-randomized: for the familiarization task, each participant was assigned to one of eight lists, making sure that no more than three items from the same condition (filler vs. experimental items) appeared in immediate succession, and that half of the participants started the task with a filler word and the other half with a target item from the interference condition. For the picture naming task, the same restrictions held. Here participants got two of eight lists, one for each block (one block consisting of two rounds), ensuring that they did not get the same list in the two blocks. For the letter search task, the order of items was semi-random, ensuring that no more than three “yes” or “no” responses followed in immediate succession.

2.3.3.2 Distractor Task – Go NoGo. To temporally separate the interference phase from the final test, following Mickan et al. (2020), participants completed a 20-minute long Go-NoGo task after interference and before the final test in Italian (based on Nigbur et al., 2011, the only difference being that stimuli remained on screen for a maximum of 1000ms rather than just 200ms). No-Go false alarm rate was on average 4% (*SD* = 5%, range = 0-24%). Since this task merely served as a filler task, we did not analyze the data any further.

2.3.3.3 Final Italian Test. Finally, to assess what interference did to participants' Italian knowledge, participants were tested again in Italian on all 70 originally learned words. Participants were asked to name all pictures twice. We chose for two rounds of naming because of possible recency of exposure differences between interfered and not interfered pictures, which the EEG is sensitive to. In ERPs, (recently) repeated words and pictures (e.g., faces) elicit attenuated N400s and enhanced LPCs compared to nonrepeated words and pictures (e.g., Bentin & McCarthy, 1994; Rugg, 1990). In oscillations, picture repetition has been found to result in a decrease in induced gamma band power (Gruber et al., 2004). While our repetition difference between conditions does not appear to be of concern for the theta band analysis, ERP signatures associated with repetition differences clearly overlap in time and are opposite in polarity to the N2 (and LPC) components that we expect as a result of our interference manipulation. Having two naming rounds should enable us to disentangle the two: repetition differences should disappear after the first round of naming, and should no longer affect the second round.

Participants were asked to name pictures in Italian to the best of their knowledge. The timings were identical to those in the picture naming tasks during the interference phase. The experimenter coded answers for correctness and in doing so initiated the next trial. There was no time limit, and next to EEG data and accuracy, (delayed) naming latencies were recorded, measuring the time from question mark presentation to speech onset. The order of presentation of the pictures was again semi-random: each participant got one of six lists from the pool of lists described for the Italian posttests at the end of each learning day. We made sure that the final test list was different from both of these posttest lists for each participant.

2.4 Behavioral Response Coding and Analysis

2.4.1 Accuracy Coding

Because the majority of errors were partial productions (a participant saying ‘albera’ rather than ‘albero’; 78% of errors; 3% of all data), participants’ Italian word productions during the final test on day 3 were coded on the phoneme level. For each production, we counted the number of correctly and incorrectly produced phonemes (see de Vos et al., 2018, and Mickan et al., 2020 for details). Incorrect productions could be either insertions, deletions or substitutions (see Levenshtein, 1966). Table 3 exemplifies the scoring procedure for the ‘albera’ example.

Table 3

Scoring example, phonetically transcribed

Target word	a	l	b	e	r	o
Participant's production	a	l	b	e	r	a
Scoring	correct	correct	correct	correct	correct	Incorrect (substitution)

'Albera' would be counted as having five correct phonemes and one incorrect phoneme. Together these two numbers (5,1) formed the basis for the dependent variable for statistical modelling. For data visualization and to provide descriptive statistics, we additionally calculated an error percentage based on these two numbers. This percentage corresponds to the number of incorrect phonemes out of the total number of phonemes (e.g., for 'albera': (1/(5+1))*100 = 16.67%).

2.4.2 Naming Latency Coding

Naming latencies were measured manually from question mark presentation until speech onset using Praat (version 5.3.78, Boersma, 2001). Note that they reflect delayed naming latencies, rather than immediate naming latencies.

2.4.3 Modelling

All behavioral data were analyzed in R (Version 3.5.1, R Core Team, 2018) using the lme4 package (version 1.1-21, Bates et al., 2015). Following de Vos et al. (2018), accuracy data were analyzed using a generalized linear mixed effects model of the binomial family, fitted by maximum likelihood estimation, using the logit link function and the optimizer 'bobyqa'. The dependent measure for this analysis was the odds of correctly producing a phoneme for a given target word. A two-column matrix with the number of correct and incorrect phonemes for each target word was passed to the model as dependent variable (this is one of multiple ways of specifying the response variable in binomial models, see also:

https://www.rdocumentation.org/packages/stats/versions/3.2.1/topics/family). We tested for main effects of Interference (two levels: no interference, interference) and Round (two levels: first round, second round), as well as for their interaction to see whether the interference effect differed in magnitude across rounds. Both fixed effects variables were effects coded (-0.5, 0.5), meaning that a negative estimate for Interference reflects lower accuracy rates for interfered compared to not interfered items, a positive estimate for Round reflects higher accuracy in round 2 than round 1, and a negative estimate for the interaction of the two would reflect a smaller interference effect in round 2 than round 1. Random effects were fitted to the maximum structure justified by the experimental design (Barr et al., 2013), which initially included random intercepts for both Subject and Item, as well as random slopes by Subject and Item for Interference and Round and their interaction. Random slopes were removed when their inclusion resulted in non-convergence to fit the maximum model justified by the data, or when they correlated with each other or their respective intercept above 0.95 to avoid over-fitting. The final models included only random intercepts for Subjects and Items as well as a random slope by Subject for Interference. All *p*-values were calculated by model comparison, using chi-square tests, omitting one factor at a time (while keeping the random effects structure constant and hence chi df = 1).

Naming latencies were analyzed using a linear mixed-effects model, fitted by restricted maximum likelihood estimation (using Satterthwaite approximation to degrees of freedom). Because we are interested in naming speed differences after the artificially introduced delay, we subtracted the 2000 ms delay from each naming latency before analysis. We then log-transformed those corrected latencies and ran the linear model on those log-transformed latencies. Fixed effects were the same as for the accuracy model and the random effects structure was also determined based on the same principles. In this model, a positive estimate for Interference reflects higher RTs for interfered than not interfered items, a negative estimate for Round reflects overall faster RTs in round 2 than 1, and a negative interaction would reflect a smaller interference effect in round 2 than 1.

For the analysis of EEG signatures during picture naming in the interference phase, we additionally calculated median splits for each participant based on their naming latencies for the interfered items during the first round of the final test in Italian. We used the naming latencies of the first round because this round reflects the cleanest measure of interference strength. This choice was further reinforced by the fact that we observed a trend towards an attenuation of the interference effect in RTs from round 1 to round 2 (see section 3.2).

2.5 EEG Recording and Analysis

2.5.1 EEG Recording

Continuous EEG was recorded from 57 active Ag-AgCl electrodes embedded in an elastic cap, following the international 10-20 system (ActiCAP 64ch Standard-2, Brain Products), as well as from an electrode placed on the forehead (serving as ground). EEG signals were referenced on-line to the left mastoid and re-referenced off-line to the averaged activity over both mastoids. Eye movements were recorded from a bipolar montage consisting of electrodes placed above and below the right eye, as well as electrodes on the left and right temples. Mouth EMG was measured with two electrodes next to the upper and lower right lip to later on be able to tell when participants talked. All data were amplified with a BrainAmp amplifier, digitized with a 500 Hz sampling rate and filtered online with a high cutoff at 125 Hz and a low cutoff at 0.016 Hz. Impedances for EEG electrodes were kept below 15 kΩ.

2.5.2 EEG Preprocessing

All off-line EEG processing was done using the Fieldtrip toolbox (Oostenveld et al., 2011) in Matlab (2018b, The Mathworks Inc.). The EEG signal was re-referenced to the average activity over both mastoids, low-pass filtered at 40 Hz, segmented into epochs from

500 ms before until 1500 ms after picture presentation, and detrended using the entire epoch. Trials containing artifacts, such as blinks or muscle activity, within the time window for analysis (-200 to 1000 ms after picture presentation) were removed. Eye blinks were identified using the EOG artifact detection function implemented in Fieldtrip. In addition, trials with amplitudes below -100 μV or above 100 μV, or peak-to-peak activity greater than 150 μV were discarded. These exclusions resulted in a total loss of 8% of the data.

2.5.2.1 ERPs. For the analysis of event-related potentials, in line with previous research, the data were furthermore baseline-corrected based on the average EEG activity in the 200 ms interval before picture presentation. We subsequently averaged EEG activity for each participant across trials for each of the interference conditions.

2.5.2.2 Oscillations. For the analysis of oscillatory power differences between conditions, we first computed time frequency representations (TFRs) of power for each of the conditions. TFRs were computed time-locked to picture presentation onset at frequencies ranging from 2 to 30 Hz, using a sliding window of three cycles, advanced in steps of 10 ms and 1 Hz. The data in each time window was multiplied with a Hanning taper, and subsequently Fourier-transformed. To test for an effect of interference condition, we subsequently calculated the difference between conditions per participant relative to the average activity in both conditions for that participant. This normalization of the condition differences made additional baseline correction unnecessary. The difference was calculated such that a positive difference reflects more power for interfered compared to not interfered words. Using cluster-based permutation tests, we compared this difference between conditions to zero (i.e., to the null hypothesis that there are no differences between conditions).

2.5.3 EEG Analysis

EEG data were assessed inferentially using nonparametric cluster-based permutation tests (Maris & Oostenveld, 2007). This method allows for the statistical comparison of multi-dimensional (M)EEG data from two conditions while controlling for multiple comparisons, which arise when comparing multiple distinct data points (i.e., time-channel and channel-time-frequency data). The method first determines spatiotemporal or spatio-spectral-temporal clusters (that is clusters of adjacent time points and sensors, or adjacent time points, sensors and frequencies) that exhibit a similar difference across conditions. It does so by means of dependent-samples t-tests at each spatiotemporal or each spatio-spectral-temporal data point, thresholded at an alpha level of .05. Spatial adjacency was defined based on a neighbourhood structure in which channels had on average 6.5 neighbours. Each observed cluster's test statistic (the sum of all t-values contributing to it) was subsequently compared to a distribution of cluster statistics obtained through 2000 Monte-Carlo permutations based on random partitions of the data. P-values of the observed clusters were calculated as the proportion of these random partitions that resulted in a larger effect (i.e., a larger cluster statistic) than the observed effect. For tests with resulting *p*-values close to the critical alpha level of .05, we reran the analysis with 5000 permutations to obtain a more reliable Monte Carlo *p*-value estimate.

Using these cluster based permutation tests, we tested for differences between interfered and not interfered items at final test in Italian, both in ERPs and in oscillations. For both analyses, we first tested for an interaction of Interference (interfered vs. not interfered words) and Round (1^{st} and 2^{nd} round of final test). To do so, and following the procedure detailed in the Fieldtrip tutorial documentation, we first calculated the averaged difference between the two interference conditions (interference – no interference) for each person and for each of the two rounds. We then statistically compared the two resulting difference structures (one for each round) via a permutation test using a dependent samples t-test. A significant difference between condition differences for the two rounds reflects a significant interaction effect. Significant interactions were followed up with separate permutation tests for each of the two rounds of the final test in Italian, whereas non-significant interactions were followed up by an analysis of both rounds of the final test combined.

For the data from the first and last rounds of the picture naming task in the interference phase, we opted to analyze the two rounds separately without conducting an interaction analysis first. Our hypothesis applied most clearly to the first round of picture naming, as explained in the Results section in more detail, and the small sample size due to the median split approach was not suited for an interaction analysis.

2.5.3.1 ERPs. We hypothesized to find differences between conditions in the amplitude of the N2 component, and hence ran targeted permutation tests in a restricted time window from 200 to 350 ms. We refrained from restricting the permutation test to frontal electrodes only, because some research has shown N2s with more posterior distributions as well (see Folstein & Van Petten, 2008; Verhoef et al., 2010). In addition to that, we also ran exploratory permutation tests for a later time window (350-1000ms), which encompasses the LPC.

2.5.3.2 Oscillations. Based on previous research, we restricted the permutation tests for the time-frequency domain to the theta frequency band (targeting 4 -7 Hz). On the basis of prior studies, we could not restrict the permutation test analysis any further and hence tested for theta differences in a window from 0 until 1000 ms after picture presentation and over the entire scalp.

3. Results

3.1 Exclusions

3.1.1 Exclusions from Accuracy Analysis

For analysis of the behavioral accuracy data during the final test in Italian, we excluded words that a participant already knew in Italian before starting the experiment (as established in the familiarization task on day 1, 1% of data), words that he/she did not manage to learn in Italian (as established in the Italian posttest on day 2, 4% of data), as well as words they did not know in English (as established in the familiarization task during interference on day 3, 2% of data). In total these exclusions resulted in 6% of data loss (*M* = 6%, *SD* = 6%, range = 0-22%), hence leaving for analysis, on average, 32 out of 35 trials per round in the interference condition and 33 trials per round in the no interference condition (the maximum per round and condition being 35).

3.1.2 Exclusions from Naming Latency Analysis

On top of the exclusions for the accuracy analysis, from the latency analysis we additionally excluded trials in which participants were unable to name a picture or named it incorrectly during the final test in Italian (i.e., errors, 4% of data). We furthermore excluded trials on which participants took multiple attempts to name a picture correctly, as well as trials on which they responded too early, that is during the 2 sec delay window (10% of data). After all these exclusions, we were left with, on average, 29 trials per round in the interference condition and 31 trials per round in the no interference condition.

3.1.3 Exclusions from EEG Analysis

For the EEG analysis, we excluded all trials that were also excluded from the accuracy analysis, as well as trials with EEG movement artifacts, as described in 'EEG preprocessing'. Artifact rejection resulted in the loss of 8% of data. After all exclusions, we had, on average, 30 and 29 trials in the interference condition in round 1 and 2 respectively (range = 24 – 35), and 31 and 30 trials in the no interference condition in rounds 1 and 2 (range = 23 – 35). Note that we did *not* discard trials based on their naming performance at final test in Italian: that is, unlike for the naming latency analysis, we included trials with errors in the EEG analyses, as well as trials in which participants took multiple attempts at naming or named a picture too early (as long as this was after the critical analysis window, i.e., after 1000 ms post picture presentation, and hence did not contaminate the EEG signal). We include those trials because the EEG analyses reflect the activity in response to stimuli and are not conditional on the final response.

The same exclusion criteria held for the analysis of the interference data. Here we were left with an average of 15 and 14 trials for the low and high RT groups in the first round of picture naming during interference, and an average of 15 and 14 trials for the same groups in the last round. Cell sizes for these comparisons are smaller than for the final test, because these comparisons rely on fewer total possible trials (i.e., max. 18 trials per median split group).

3.2 Behavioral Results

3.2.1 Learning Performance in Italian

After the first Italian learning session on day 1, participants had learned on average 46 out of the 70 words (M = 46.48, SD = 10.85, range = 26-69). Learning success on day 1 was comparable between the two interference conditions (Interference: M = 66%, SD = 17%, range = 26-97%; No interference: M = 67%, SD = 16%, range = 40-100%). After the second Italian learning session on day 2, participants had learned on average 67 out of 70 words (M = 67.37; SD = 3.56. range = 56-70). Learning success was again equal for both interference conditions (Interference condition: M = 96%, SD = 5%, range = 83-100%; No interference condition: M=96%, SD = 6%, range = 77-100%), and overall very high.

3.2.2 Retrieval Performance in Italian after Interference on Day 3

3.2.2.1 Naming Accuracy. Mean error rates for the interfered and the not-interfered items during final test in Italian are shown in Figure 2 and the corresponding model output is reported in Table 4. We observed a main effect of Interference in the predicted direction: participants made more phoneme production errors on interfered compared to not-interfered words. In model terms, this main effect is reflected in a negative estimate, because we model accuracy rather than errors, and phoneme production accuracy for target words is lower in the interference condition than in the no interference condition. There was also a main effect of Round, such that participants improved and made less errors overall from round 1 to round 2. Round, however, did not modulate the main effect of interference. The interference effect in accuracy / error rates was thus stable across the two rounds of the final test.

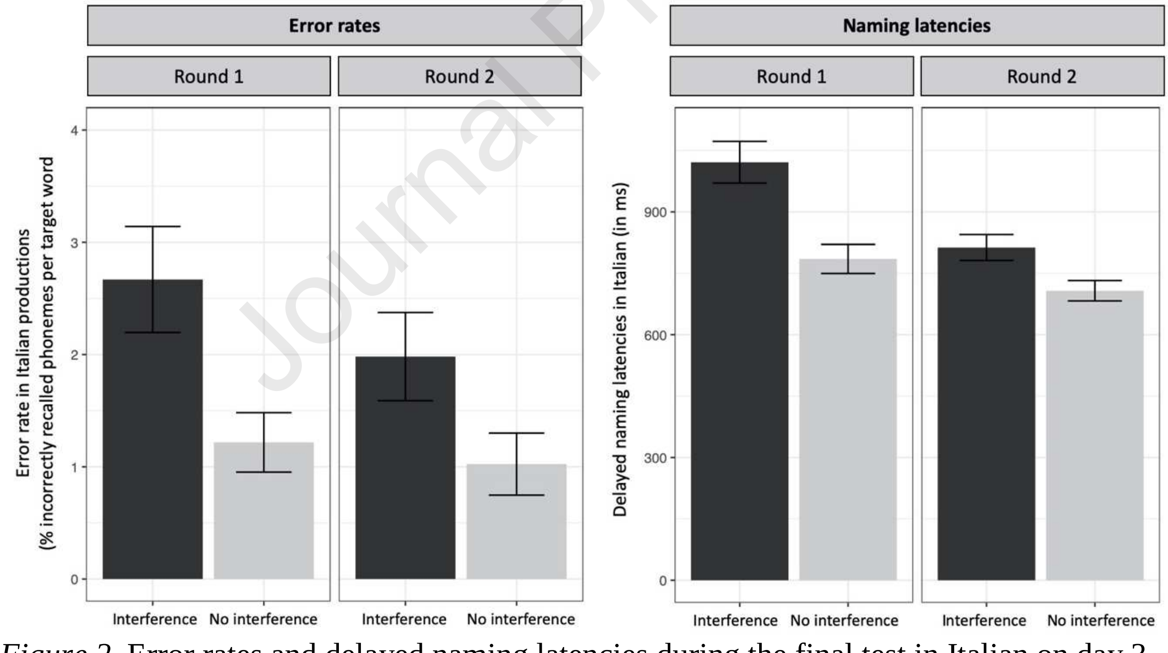


Figure 2. Error rates and delayed naming latencies during the final test in Italian on day 3. Error rates are expressed in the number of incorrectly produced phonemes per target word, and naming latencies reflect the time it took participants to name a picture after a 2 sec delay period.

Table 4

Mixed effect model output for naming accuracy in the final Italian test on day 3

Fixed effects	Estimate	*SE*	*z*	$p(\chi^2)$
Intercept	5.91	0.36	16.44	**<.001**
Interference	-0.77	0.31	-2.44	**.026**
Round	0.27	0.11	2.54	**.015**
Interference * Round	0.20	0.21	0.91	.389
Random effects	Groups	Var	*SD*	Corr
Item	Intercept	2.91	1.71	
Subject	Intercept	1.75	1.32	
	Interference	1.62	1.27	0.08

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

3.2.2.2 Naming Latencies. Mean naming latencies for interfered and not interfered items are shown in the right panel of Figure 2, and corresponding model outcomes in Table 5. We observed a main effect of Interference, such that interfered words took participants longer to recall than not-interfered words. We also found a main effect of Round: participants were overall faster in round 2 compared to round 1 of the Italian final test. The interference effect was numerically bigger in the first round, the corresponding interaction term, however, did not reach statistical significance, indicating that the interference effect was present in both rounds and did not differ significantly in magnitude between rounds. Follow-up models for each round separately confirm this (round 1: β = 0.15, t = 5.33, $p(\chi^2) < .001$; round 2: β = 0.08, t = 2.96, $p(\chi^2)$ = .006).

Table 5

Mixed effect model output for log-transformed naming latencies in the final Italian test on day 3

Fixed effects	Estimate	*SE*	*t*	$p(\chi^2)$
Intercept	6.47	0.05	125.51	**<.001**
Interference	0.12	0.02	4.94	**<.001**
Round	-0.08	0.02	-4.57	**<.001**
Interference * Round	-0.07	0.04	-1.92	.055
Random effects	Groups	Var	*SD*	Corr
Item	Intercept	0.03	0.18	
Subject	Intercept	0.06	0.24	
	Interference	0.01	0.08	0.77

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

3.3 EEG Results

3.3.1 EEG – Final Test in Italian

Grand-averaged ERPs for the interfered and not interfered words during rounds 1 and 2 of the final test in Italian are shown in Figure 3. A time-frequency representation of the difference in induced activity between interfered and not interfered words is shown in Figure 4.

3.3.1.1 ERPs – N2 Time Window (200-350 ms). An initial permutation test revealed a significant interaction between Interference and Round (p = .002). This interaction was most prominent in an interval from 212 to 350 ms. Subsequent separate permutation tests for each of the two rounds of the final test revealed a large positivity for interfered compared to not interfered words in the first round (p = .001). This effect was present throughout almost the entire analysis time window, but most prominent between 204 and 350 ms and over centro-posterior electrodes. Visual inspection reveals that this positivity for interfered items is best described as an attenuated negativity for interfered compared to not interfered items (see

Figure 3). The direction of the effect and its scalp topography suggest that this component reflects the beginning of an attenuated N400 for more recently seen pictures (i.e., the interfered items) compared to less recently seen pictures (i.e., the not-interfered items). A follow-up analysis on a time window encompassing the classical N400 effect (200-500 ms) confirms this: the permutation test again revealed a significant positive shift (or in other words, a less negative shift) for interfered compared to not interfered items in this window ($p = .002$), which was most prominent over centro-posterior electrode sites and from 204-428 ms.

In the second round, we instead observed the expected N2 modulation. The permutation test revealed a larger negativity for interfered compared to not-interfered items ($p = .019$). This difference between conditions was most pronounced in a time window from 218 to 316 ms and over frontal electrodes, which coincides well with the typical time course and topography of the N2. The ERP signatures in this early time window thus reverse from round 1 to round 2. The N2 effect in the second round confirms our hypothesis and the reversal of signatures suggests that recency differences between items were successfully eliminated after the first round of naming.

3.3.1.2 ERPs – Later Time Window (350-1000 ms). The interaction term between Interference and Round from 350 to 1000 ms post picture presentation did not reach statistical significance ($p = .061$). A follow-up permutation test over both rounds of the picture naming test together revealed a wide-spread negative cluster for interfered compared to not-interfered items ($p = .007$). Visual inspection of the grand average revealed that this cluster reflects a late positive component (LPC), that is attenuated for the interfered items as compared to the not-interfered items, most pronounced from 428 to 636 ms. This LPC is present in both rounds (though see Fig. 3 for grand averages and cluster plots for each round separately).

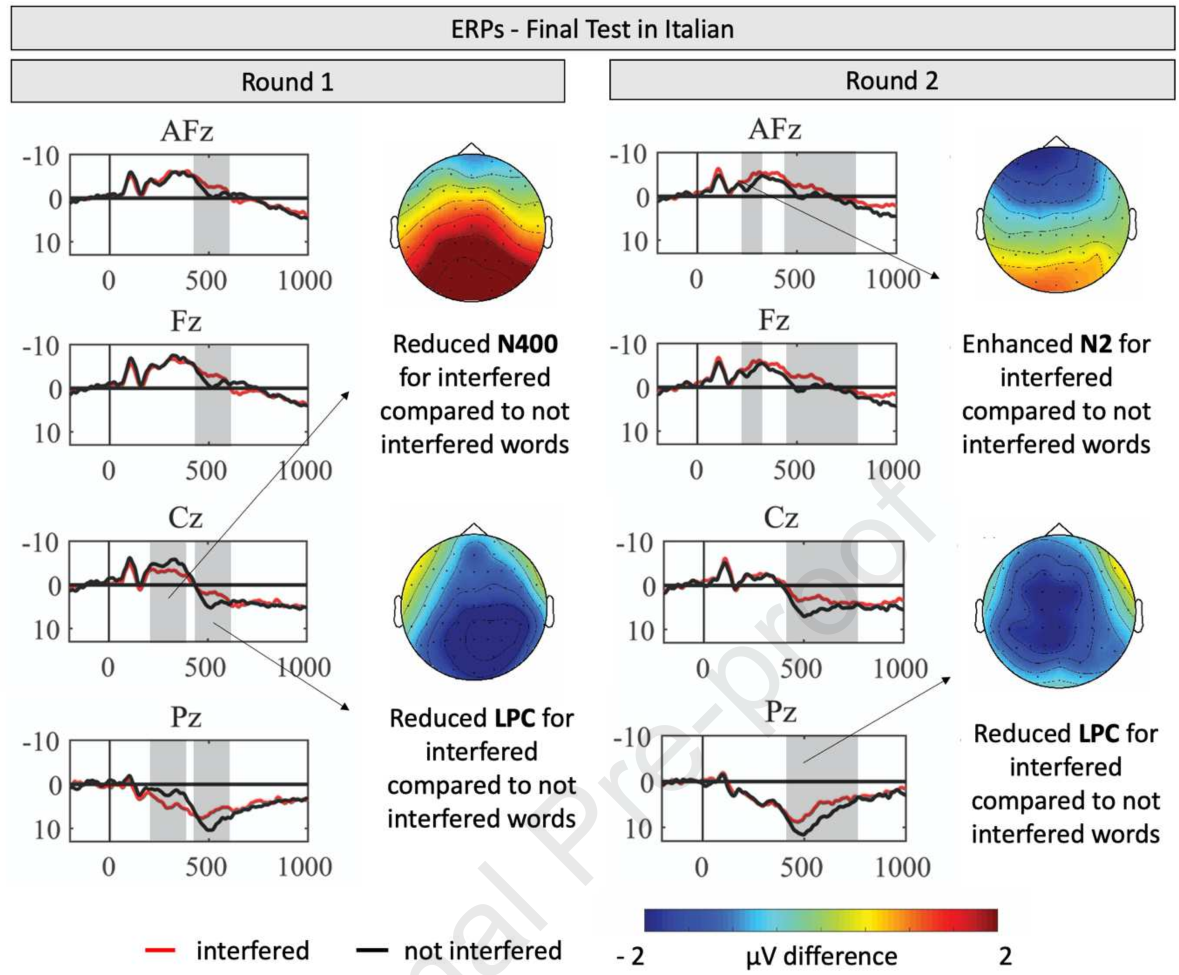


Figure 3. Grand-averaged ERP waveforms for interfered and not-interfered items during rounds 1 and 2 of the final Italian picture naming test. Significant clusters revealed by the permutation tests are marked in grey. For each cluster a topographic plot is included. Colors indicate the amplitude difference (in µV) between interfered and not-interfered items, such that shades of red reflect more positive going ERPs for the interfered compared to the not-interfered items, and shades of blue reflect more negative going ERPs for interfered items.

3.3.1.3 Oscillations – Theta Band (4-7Hz). In the time-frequency domain, there was no significant interaction between Round and Interference (p = 1). A follow-up permutation test of the data collapsed over both rounds of the final naming test revealed a large cluster in the theta frequency band (p = .004). Retrieval of interfered items thus resulted in more induced theta activity than retrieval of not-interfered items, which we observed most prominently in a time interval of 510-1000 ms and distributed over the entire scalp.

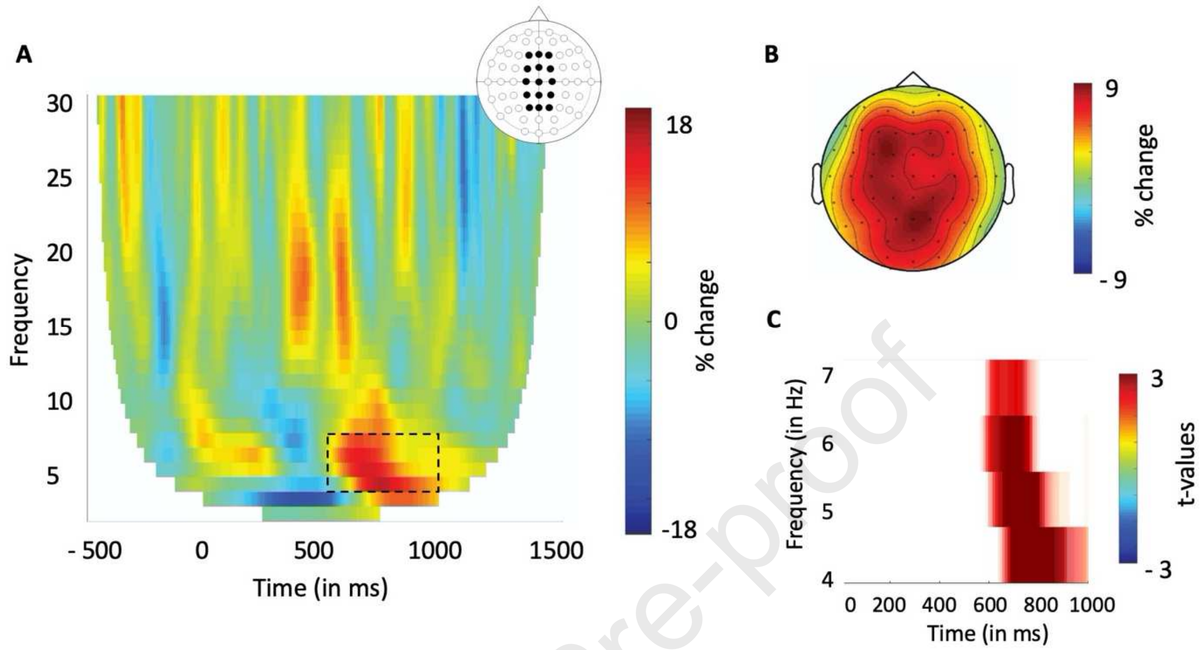


Figure 4. **A.** Time-frequency representation of power differences between interfered and not-interfered items, averaged over a representative sample of channels involved in the cluster revealed by the permutation test (see black dots in the topoplot in the right upper corner: Fz, F1, F2, Cz, C1, C2, FCz, FC1, FC2, CPz, Cp1, Cp2, Pz, P1, P2). Shades of red reflect more theta for interfered compared to not interfered items. Power differences were calculated relative to the average activity in both conditions, and thus reflect a percent power change. Dashed lines reflect the significant cluster. **B.** Scalp distribution of power changes for the interference condition minus the no-interference condition (relative to the average activity in both condition), averaged from 510-1000 ms and for frequencies between 4-7 Hz. **C.** Statistical map for the theta effect in time and frequency, averaged over the same channels as in A. Colors reflect t-values.

3.3.2 EEG – Interference Phase in English – Interfered Items Only

To test whether activity during the interference phase was directly related to retrieval performance at final test, we analyzed the interference phase data conditional on participants’ naming latencies at final test in Italian. Based on a median split, we divided each participant's interfered items into those that took participants long to recall at final test in Italian and those that took them relatively less long to recall. The inhibitory control account of forgetting would attribute such retrieval difficulty discrepancies to differences in inhibition during the interference phase: Italian labels that are more difficult to recall at final test must have been inhibited more during retrieval of their English translation equivalents in the interference phase. If competition and inhibition during the interference phase are indeed responsible for the retrieval difficulty differences between items at final test, we should thus observe a higher amplitude N2 and more theta power for items that are slow to retrieve at final test (high interference group) compared to items that are fast to retrieve at final test (low interference group). This hypothesis concerns most directly the first round of picture naming in the interference phase. We speculate though that by the last round, differences between the two conditions disappear. To that end, we analyzed grand averages for both the first and fourth (i.e., last) round of picture naming during interference. Note that because the second block of picture naming was preceded by two rounds of phoneme monitoring in English, round four of picture naming corresponds to round 6 of the interference phase overall. Rounds 1 and 4 are thus separated by 5 intermittent retrievals rather than just 3 (as their names might suggest). Note also that because of these intermittent retrievals and the nature of the comparison (low vs. high interference), round 4 of picture naming during interference is not equivalent to round 2 of picture naming at final test in Italian (where interference is compared with no interference at all). Grand averages and topoplots contrasting the two median split interference groups are shown in Figure 5.

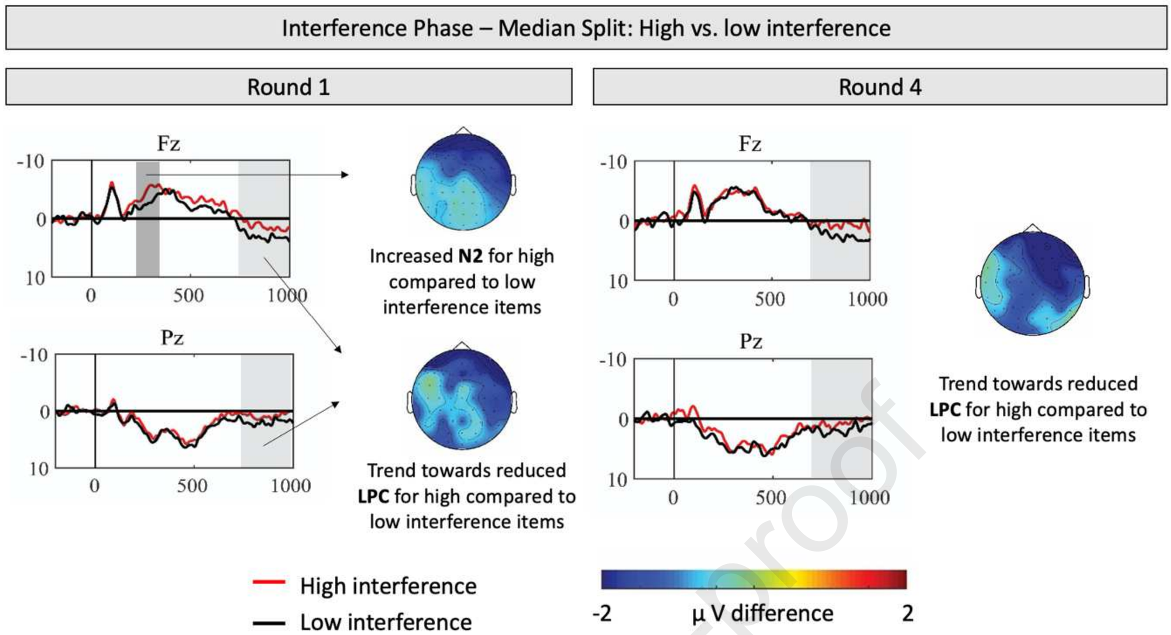


Figure 5. Grand-averaged ERPs for items from the high and low interference groups (as determined through a median split of naming latencies from the final Italian test) from the English interference naming task, rounds 1 and 4. Topographies of significant effects and trends in the data are displayed to the right of their respective grand averages. Colors indicate the amplitude difference (in µV) between conditions, such that shades of blue reflect more negative going waveforms for highly interfered items compared to less interfered items.

3.3.2.1 ERPs – N2 Time Window (200-350 ms). In the first round of picture naming during interference, we indeed observed a larger N2 for highly interfered items (i.e., items that later took relative long to produce in the final Italian test) compared to less interfered items (p = .049). The difference between conditions was most pronounced over frontal electrodes and between 218 and 350 ms post picture presentation. In the last round of picture naming, this N2 was no longer present (i.e., no significant clusters, ps = 1).

3.3.2.2 ERPs – Later Time Window (350-1000 ms). In the later time window, visual inspection suggests that there is a small late positive shift for high compared to low interference items both in the first and the fourth round of picture naming during interference. These differences, however, were not statistically robust (1st round: *p* = .066, differences most pronounced between 730-1000 ms; last round: *p* = .051, differences most pronounced between 728-1000 ms). These positive components differ from the LPC reported in the final test both in their temporal as well as their spatial distribution.

3.3.2.3 Oscillations – Theta Band (4-7 Hz). There were no significant differences between high and low interference items in the time-frequency representations of either of the two rounds of picture naming in the interference phase.

4. Discussion

The present study aimed at unravelling the neural correlates of foreign language attrition. Previous behavioral studies postulated that foreign language forgetting is the consequence of competition and inhibition between translation equivalents (Mickan et al., 2020). Here, we asked whether we can track those processes on the neural level. To that end, participants first learned a set of new L3 Italian words over two consecutive days. On a third day, we interfered with their knowledge of these recently learned words by having them repeatedly retrieve half of the words in L2 English. Finally, we assessed the effect of this interference phase in a final recall test on all originally learned Italian words, also on day 3. Next to asking whether we can see neural evidence for competition and inhibition at final test, we also asked whether behavioral performance at final test can be related to the degree to which these processes are recruited during interference.

Behaviorally, we replicated Mickan et al. (2020): participants were slower and less accurate in recalling Italian words that had been interfered with (i.e., named in English) than words that had not. In the EEG, these interference effects were accompanied by more theta power, an enhanced N2 and a reduced LPC for interfered compared to not interfered items. Moreover, differences in performance at final test went along with amplitude differences in the N2 component during the interference phase: we report an enhanced N2 for items that took participants long to recall at final test compared to items that were easier to retrieve and hence interfered with less successfully. Together, these findings establish the N2, the LPC and oscillatory power in the theta band as neural correlates of foreign language attrition.

4.1 Behavioral Evidence for Between-Language Competition in FL Attrition

Replicating the behavioral interference effects reported in previous lab-based language attrition studies (e.g., Mickan et al., 2020) with a new language combination and a larger set of to be learned words confirms the robustness of these effects: foreign language forgetting can be the consequence of the recent use of another foreign language. These interference effects occur despite the fact that the learning phase in our experiment was spread over two days and even though the reaction times at final test were measured after a delay rather than immediately after picture presentation.[2] What is more, unlike most previous studies on bilingual language production, our results demonstrate that between-language interference unfolds not just globally (i.e., between entire languages; e.g., Costa & Santestban, 2004; Kreiner & Degani, 2015; Meuter & Allport, 1999), but also locally, that is, on the item-level between translation equivalents. Local interference effects have only rarely been documented. In fact, studies looking at item-specific interference effects have

[2] In an almost identical design but *without* a production delay at final test, Mickan et al. (2020) report average naming latencies of roughly 2200 and 1700 ms for interfered and not interfered words respectively. Naming latencies for remembering words in a foreign language a day after having learned them (and after an English interference phase) appear to be fairly long even without an enforced production delay. The delay of two seconds we introduced here was thus unlikely to wash out retrieval speed differences between interfered and not interfered words.

sometimes reported the opposite, namely translation facilitation (e.g., Branzi et al., 2014; Misra et al., 2012; Wodniecka et al., 2020).

Evidence for translation facilitation comes from blocked language switching studies, in which participants are first asked to name a set of pictures in, for example, L1, followed by a block of naming of (partially) the same pictures in L2 (e.g., Branzi et al., 2014; Misra et al., 2012; Wodniecka et al., 2020, instead looked at L1-after-L2 naming). While the blocked design of these experiments is reminiscent of the set-up of the current study, the comparisons they make are fundamentally different from ours. To investigate item-specific interference effects, Branzi et al. (2014), for example, studied how naming in L2 after having named the same pictures in L1 compares to naming in L2 after *no* prior naming (i.e., naming new pictures in L2). Not surprisingly, they report facilitation for naming a picture in L2 if the same picture had previously been named in L1 compared to when it had not been named before at all (see also Misra et al., 2012, and for comparable effects when L1 naming followed L2 naming see Wodniecka et al., 2020, but see Experiment 2 in van Assche et al., 2013, for a null effect). The conceptual and visual facilitation that picture repetition had on L2 naming in these studies very likely washed out any potential interference effects from prior L1 naming. In contrast, in our study, participants were familiarized with pictures in both the interference and no interference conditions through the two-day learning session in Italian (on average 15 exposure per item). The additional exposures to pictures in the interference condition during the English interference phase most likely induced only minimal visual and conceptual priming differences between conditions (the benefit of picture repetition decreases with every extra picture repetition; Gollan et al., 2005; Griffin & Bock, 1998). We speculate that minimizing such facilitation effects is what enabled us to observe inhibitory rather than facilitatory effects between translation equivalents. Other studies that do include a familiarization (or initial learning or test) phase in an otherwise similar blocked design also report item-specific interference effects just like we do, reinforcing this explanation (see Degani et al., 2019, for reduced accessibility of L1 words after exposure to L2 translations; and Bailey & Newman, 2018, for reduced accessibility of L2 after exposure to L1). Overall, our behavioral results thus differ in a number of ways from previous research on so-called language “after-effects” and advance our understanding of how two foreign languages interact with one another.

4.2. Neural signatures of between-language competition and interference

The EEG signatures of between-language competition and interference that accompany the behavioral effects we report resemble those reported in various other strands of literature, including research on bilingual language production and forgetting more generally. Departing from how they are typically interpreted in these other areas, our EEG results provide converging evidence for the assumption that (foreign) language attrition can be the consequence of competition and interference from the more recent use of other languages.

4.2.1 The N2 as a Marker for Interference-Induced Foreign Language Attrition

The frontal N2 component that we report for interfered compared to not interfered items in the second round of the final test resembles the N2 that is often found in language switching studies. In those studies, participants typically alternate between naming pictures in L1 and L2, and the N2 is found to be strongest on switch compared to repeat trials (particularly when a switch is made from L1 to L2; Jackson et al., 2001; Zheng et al., 2020). In line with reports of the N2 as a marker of response conflict and inhibition in non-linguistic tasks (e.g., Folstein & van Petten, 2008), language switching studies typically interpret their results as evidence for interference from and inhibition of a non-target language (e.g., the L1 when switching to naming pictures in L2; see Kroll et al., 2008, for a review). Observing a comparable N2 for interfered items at final test is thus compatible with the idea that between-language interference is (at least partially) responsible for the behavioral forgetting effects measured at final test. Specifically, it is in line with the proposal from Mickan et al. (2020) that retrieval of interfered L3 words is hindered by competition from the recently practiced L2 words and that this interference is not (or much less) present for L3 words whose L2 translations were not recently retrieved. Whether the N2 reflects only the presence of this response conflict (i.e., interference between English and Italian labels), or in fact the active inhibition of the English competitors to allow for successful retrieval of the Italian words, is unclear. In fact, it might also simply reflect retrieval difficulty, that is the consequence of increased interference and competition rather than the presence of these processes per se (see Wodniecka et al., 2020, for this proposal).[3] On any of these accounts, however, the N2 provides corroborating evidence in favor of the idea that language forgetting can be caused through interference from recently retrieved translation equivalents.

Our N2 is comparable to the switching N2 both in terms of latency (200-350ms post stimulus presentation) and scalp topography (fronto-central). This is interesting and, in fact, not trivial, because our study differs from mixed language switching studies in a number of ways. As explained in the Introduction, these differences include the timing of the switch (immediate vs. delayed), the level at which interference / inhibition is thought to act (language global vs. local, item-specific), and the languages involved (L1/L2 vs. L2/L3). We are only aware of three EEG studies that have addressed long-term switch effects and that tested item-specific switching on top of global switch effects (Branzi et al., 2014; Misra et al.,

[3] Given that retrieval difficulty is the direct consequence of increased interference and competition, these two interpretations are impossible to tease apart. It is worth noting though that the mainstream interpretation of the N2 to date invokes interference and inhibition rather than retrieval effort, which is why we stick to the former in the remainder of the paper.

2012; Wodniecka et al., 2020), but, for reasons explained in section 4.1, they are not comparable with our design or with the studies by Mickan et al. (2020) and Bailey and Newman (2018). The current study thus differs from both mixed and blocked language switching studies in important ways. That we nevertheless report a comparable N2 effect is in line with the idea that similar inhibition and interference mechanisms might be at work in language switching and L2-induced L3 attrition. Just as global switching from naming pictures in L1 to naming pictures in L2 invokes an N2, so does the retrieval of words in Italian after a remote block of naming the same items in English.

4.2.2 Oscillatory Theta Power as an Index of Between-Language Competition

In the frequency domain, we report more theta power for interfered compared to not interfered words at final test in Italian. Though different in terms of scalp topography, our theta effect fits with reports of interference-induced theta activity in other domains, such as, for instance, the non-linguistic cognitive control literature. In a go/no-go task, for example, mid-frontal theta power is typically higher on no-go trials, where the tendency to press a button needs to be suppressed, compared to go trials (e.g., Nigbur et al., 2011). Very similar to the N2, theta is hence understood to index the presence of a response conflict and possibly the recruitment of cognitive control processes to overcome this conflict. Next to the cognitive control literature, memory research on so-called retrieval-induced forgetting (RIF) effects has also consistently reported modulations in the theta band. These studies reported higher mid-frontal and left parietal theta power in competitive compared to uncompetitive retrieval situations, suggesting that theta indexes the amount of competition and thus interference that is encountered during item recall (e.g., Staudigl et al., 2010; Hanslmayr et al., 2010). Our theta effect is not restricted to mid-frontal or left-parietal electrode sites, and is instead more wide-spread. This topography difference is most likely attributable to differences in stimuli and task design between our experiment and the theta studies in other domains. Competition from translation equivalents and the suppression of a non-target language word likely requires a different kind of control than the suppression of a 'Go' response in a no-go trial or the suppression of semantic competitors in RIF paradigms. Remember also that some of the RIF studies compare two different tasks (e.g., active retrieval vs. passive restudy in Staudigl et al., 2010) and that the scalp topography of theta activity reported in these studies might thus also partially reflect differences in task design between the two conditions rather than interference alone, making it difficult to compare to our theta effect.

Regardless of the topography differences, we think that it is justified to conclude that the theta effect in our study reflects interference of a non-target language (i.e., English) during productive recall of words in a target language (i.e., Italian). Just as the N2 discussed above, the theta effect at final test thus corroborates the idea that between-language competition is at least part of the reason for why interfered Italian words at final test are less well recalled.[4] To our knowledge, we are the first to provide evidence for increased theta power as a marker of between-language interference.

4.2.3 The Consequences of Language Interference – the LPC

In the final test, next to the N2 and theta effects, we additionally observed a late positive component (LPC), reduced in magnitude for interfered compared to not interfered

[4] Note that we do not claim that theta and N2 reflect the exact same underlying processes. In fact, as we discuss in section 4.4, they dissociate in some experimental phases. Future research will be necessary to disentangle to what extent they reflect the same underlying cognitive mechanisms. Note also that as with the N2, it is possible to interpret theta as a marker of retrieval effort (see footnote 3). Since we have not come across this interpretation of theta in the cognitive control or RIF literature though, and since it is no more plausible than the interference/competition interpretation, we stick with the latter account.

items in both rounds at final test. Both in terms of its central scalp distribution and latency (roughly 400-600ms post stimulus onset), this signature is reminiscent of a similar late positive component in the memory literature. As explained in the Introduction, the ‘memory’ LPC is most typically found in studies on recognition memory, where it is stronger at retrieval for previously studied (‘old’) compared to previously unstudied (‘new’) items, and especially for items for which participants additionally make correct as compared to incorrect source judgments (i.e., recalling details of the original learning context; Rugg et al., 1995; Wilding, 2000). Its amplitude has furthermore been found to vary with decision certainty, such that it appears to be larger for items that people report to confidently remember as compared to items for which people only report a vague sense of familiarity (Smith, 1993). Given the conditions that elicit this component, the LPC is generally understood as a marker of conscious recollection success, and possibly an index of the quality of the information that is retrieved from episodic memory.

Though not specifically predicted, our finding of an enhanced LPC for not-interfered compared to interfered items fits well with this recollection-success interpretation. Memory representations of Italian labels in the no interference condition have not been interfered with and so retrieval for those items is easier, faster and ultimately more successful (as also seen in reaction times and error rates) than for interfered items. It thus seems plausible that the LPC in our study indexes retrieval success in Italian. Note that one could have also predicted the opposite pattern: a larger LPC for the interfered items because their corresponding pictures have been repeated more recently (Bentin & McCarthy, 1994). That this was not the case reinforces the interpretation that the LPC in our study indexes recollection processes specific to the Italian words, and not their associated concepts.

In the language domain, LPC effects have been found to index lexicality and conscious semantic access. Bakker et al. (2015), for example, reported a reduced LPC for newly learned words (in L1) compared to existing words and partial evidence for an increase in the magnitude of the LPC with consolidation of these novel words. Their LPC effect, however, had a fronto-central scalp distribution and was furthermore elicited under very different task demands (semantic relatedness judgments between the words and unrelated primes), and is hence difficult to compare directly to our findings. Even though the comparison is not straight-forward, if our LPC were to index degree of lexicality, this would mean that words in the interference condition, despite having been learned to the same criterion as not-interfered words, lack behind in lexicalization, or that their lexical representations have undergone erosion due to interference. In a follow-up experiment, it would be interesting to establish degree of lexicality (i.e., LPC amplitude) prior to interference, to see exactly what changes interference brings about, and to be able to tell whether interfered items decrease in lexicality (i.e., decrease in LPC magnitude) due to interference or simply stagnate, compared to not-interfered items (i.e., LPC amplitude increases for not-interfered items and remains the same for interfered items).

Curiously, some of the mixed language-switching studies described earlier tend to report an LPC opposite to that in our study (i.e., larger for switch compared to repeat trials, essentially a continuation of the earlier N2, e.g., Jackson et al., 2001). Not all language switching studies report an LPC though, making it unclear what the precise conditions for its emergence are. Most likely, the switching LPC reflects different processes than the LPC we report here and future research will be necessary to fully understand its functional significance in multilingual language production. Based on the present results, and the available evidence from other strands of research, we conclude that the LPC is a marker for retrieval success and as such reflects the consequence of between-language interference, namely reduced accessibility to interfered compared to not interfered Italian labels.

4.3 Disentangling Recency from Interference

One aspect of the final test that warrants discussion is the fact that we observed the predicted N2 modulation only in the second round of the final test, whereas we did find effects in theta power and the LPC in both rounds. In place of the N2, we observed a reduced (rather than enhanced) negativity for interfered compared to not interfered items in the first round of the final test, which we interpreted as an attenuated N400 based on its latency and topography. This N400 most likely reflects recency differences between items in the two conditions. Though equally familiar initially, the pictures corresponding to the interfered items were seen more recently than those of the not-interfered items, and hence were less surprising and easier to process, resulting in an attenuated N400 (Bentin & McCarthy, 1994). Differences between conditions caused by recency appear to be much stronger than differences due to interference and so the N400 (larger in amplitude for *not* interfered items) overwrote the N2 (larger in amplitude for interfered items) in the first round. By round two, recency differences between items had disappeared, enabling us to observe the predicted interference-related N2. In contrast, neither the LPC nor theta power appear to be influenced by such recency differences. In the frequency domain, previous literature only implicated the gamma frequency range in picture repetition (Gruber et al., 2004). The LPC, in turn, has been found to be sensitive to picture repetition, yet in the opposite way, being larger for repeated (i.e., interfered items in our study) compared to not repeated items (e.g., Bentin & McCarthy, 1994). The processes that our LPC effect reflects (i.e., recollection success for Italian labels) appear to have been stronger than item differences due to picture repetition.

While this confound is unfortunate, we would like to stress that recency differences are inherent to the design of our study. Eliminating them would require inclusion of the no-interference items in the interference phase, in a task that does not require competitive retrieval of these words, but nevertheless exposes participants to their images. One could argue that we could have used a simple passive exposure task, akin to the EEG RIF studies mentioned earlier. However, given that our stimuli are meaningful words, relevant not only within the context of the experiment itself, it is very possible that even in a passive exposure condition (or in fact in any task), participants would covertly retrieve the words (in whatever language). Such word retrieval would have interfered with our experimental manipulation in that the words from the no-interference condition would then also have received interference.

To weaken recency differences, future studies could use different pictures in each experimental phase: all pictures would be equally new at final test then and hence differences in ease of visual recognition would no longer contaminate the signal. Note though that items in the interference phase would still be *conceptually* more recent and might thus still be easier to access even with a different set of pictures. The latter risks and considerations are why we instead stayed with the paradigm established by Mickan et al. (2020).

4.4 Linking Activity During Interference to Later Forgetting

So far, we have looked at EEG activity during final recall of Italian items and found evidence for competition and interference at that moment (theta and N2) as well as the immediate consequences of this interference for recall success (LPC). While competition and interference at final test suffice to explain the observed behavioral forgetting effects, interference-driven (language) forgetting is typically assumed to already be induced during the preceding interference phase (Anderson, 2003; Mickan et al., 2020). Studies on the neural correlates of retrieval-induced forgetting support this claim (e.g., Johansson et al., 2007; Hanslmayr et al., 2010). Staudigl et al. (2010), for example, found that participants who showed the greatest decrease in theta activity over multiple rounds of competitive retrieval (in the interference phase) also forgot more of the very competitors that caused the competition during retrieval. Staudigl and colleagues interpret the competition reduction that takes place across subsequent rounds of retrieval to reflect the amount of inhibition that is applied to competitors. The more inhibition is applied, the more troublesome retrieval is for those competitors at subsequent final test, and hence the larger the forgetting effect.

Here, we asked whether a similar relationship between activity during the interference phase and final test also holds for the language case. Our median split analysis of the interference phase data reflects a first step towards understanding the temporal dynamics of interference-induced foreign language attrition. We split each participant's items into high and low interference items depending on how fast they were recalled at final test. Items that took a participant relatively long to recall at final test must have been interfered with more than items that were faster to recall at final test. The former should hence show more evidence for interference (and possibly inhibition) during the interference phase than the latter, if there is a direct relationship between the two experimental phases. While we did not observe a modulation of theta power during the interference phase, we did find differences between the two types of items in the amplitude of the N2 component. In the first round of picture naming during the interference phase, we observed a higher N2 amplitude during English retrieval of items that were subsequently more difficult to retrieve in Italian than items that were relatively easy to retrieve at final test. There is thus indeed a quantifiable relationship between activity during the interference phase and later retrieval ease. Assuming that the N2 reflects the presence of interference from response alternatives (i.e., Italian labels during English picture naming) and possibly the need for inhibition of those competing responses for successful retrieval of the target response (i.e., the English label), the current pattern of results suggests that the extent to which Italian labels interfered and/or were inhibited is directly related to how well they were recalled at final test. The behavioral interference effects are thus not only the result of competition at final test, but are already set in motion during the preceding interference phase.

Interestingly, in the last round of picture naming during the interference phase, the N2 was no longer enhanced for highly interfered as compared to less interfered items, suggesting that retrieval differences at final test are induced at the beginning of the interference phase rather than later on. After multiple rounds of retrieval in English, the Italian translations in the high interference group no longer interfered more and no longer needed extra inhibition than items in the low interference condition. It should be noted though that this decrease was only descriptively observed in the current study. The small sample size did not allow for a statistical comparison of the two rounds of picture naming in the interference phase (i.e., no interaction analysis with round was possible).

We encourage future research that follows up on our interference phase analysis, not only to replicate the N2 findings, but also to better understand why neither theta power nor the LPC amplitude reliably distinguished later well and less well recalled items. As already noted, the interference phase analysis is based on a relatively small number of trials per condition (15 trials on average) and so it is possible that we simply did not have enough power to reliably detect theta power and LPC amplitude differences. A follow-up study with more items, and possibly without a no-interference condition (allowing for all 70 learned Italian items to be part of the interference phase) would help explain the current pattern of results.

4.5 A Note on Language Strength and How It Relates to Interference Magnitude

Overall, both the behavioral and the EEG results support the conclusion that using an already known foreign language can hamper subsequent access to a just recently learned other foreign language. More specifically, we have documented interference from a relatively stronger foreign language (L2 English) on a (supposedly) weaker foreign language (L3 Italian). Interestingly, the majority of previous research on bilingual language production focused on interactions between L1 and L2 and mostly found the stronger L1 to be negatively affected by a previous naming block in a weaker L2. Speaking in a weaker L2, in turn, has often not been found to be (negatively) affected by a prior block of naming in the stronger L1. In section 4.1, we already discussed that Branzi et al. (2014) and Misra et al. (2012) found that L1 naming had a positive rather than negative effect on later naming of the same pictures in L2. For the opposite block order, when naming in L1 was preceded by naming of the same pictures in L2, no such facilitation was observed. They interpreted this difference as evidence that L2 naming requires more inhibition of L1 and hence that a prior naming block in L2 induces more interference for subsequent L1 naming than a prior naming block in L1 does for subsequent naming in L2 (but see Wodniecka et al., 2020, for facilitation effects also for L1-after-L2 naming). Moving away from item-specific interference effects, global inhibition/interference effects also appear to be stronger when L1 naming follows L2 naming rather than the other way around (e.g., Branzi et al., 2014; see also the switch-cost asymmetry in mixed language switching studies: Bobb & Wodniecka, 2013; Meuter & Allport, 1999).

Accordingly, it has been proposed that between-language interference and inhibition only arise when speaking in a relatively weak language (i.e., in L2, when L1 needs to be inhibited), but not while speaking in L1, and hence that speaking a stronger language (e.g., L2 in our study) should not hamper the subsequent retrieval of weaker languages (e.g., L3 in our study). Our results, however, suggest that this can be the case. Our interference effects can only be explained by assuming that the recently learned, supposedly still weak L3 Italian words *did* interfere with their (supposedly stronger) L2 English translation equivalents during the interference phase and that because of that they had to be inhibited (or conversely their English equivalents had to be boosted), resulting in later retrieval difficulties in Italian at final test. While our findings appear to be at odds with some previous studies, we are not the first to observe competition effects from a weaker on a stronger language (see Klaus et al., 2018; Lemhöfer et al., 2018), or to observe a negative “after-effect” of exposure to a relatively stronger language on subsequent retrieval of a weaker language (e.g., Bailey & Newman, 2018; Kreiner & Degani, 2015). More importantly though, interference strength is likely not only affected by language strength, but also by recency of exposure differences between items (see the retroactive interference literature, e.g., Wixted, 2004). Since we were not interested in relative strength differences between languages, these two aspects are confounded in our study. The Italian words were learned in an extensive learning session spread out over two days. Although they were thus new and still weak, they were also very fresh in our participants' memory. It is consequently unclear whether our results are in fact directly in conflict with earlier studies, where strength and recency were not confounded in the same way. We hope future research will clarify the relative contributions of strength and recency on interference magnitude.

4.6 Summary

The current study established the N2, the LPC and oscillatory power in the theta band as neural markers of foreign language attrition. Their presence at final test and (at least partially) during the interference phase supports the idea that foreign language forgetting is the result of competition dynamics between translation equivalents in multiple languages. At final test in Italian, oscillatory power in the theta band and the N2 component of the event-related potential reflected interference from (and possibly inhibition of) the recently practiced English translation equivalents. The LPC, in turn, based on its occurrence in the memory literature, most likely reflected the consequences of this competition between English and Italian labels and indexed the reduced accessibility to interfered compared to not interfered Italian labels. Finally, we were able to link activity during the preceding English interference phase to later retrieval speed in Italian: an enhanced N2 for items that were later most difficult to retrieve is in line with the idea that competition and inhibition during the interference phase are causally related to later retrieval ability at final test. Taken together, our results provide the first converging neural evidence for the idea that foreign language attrition can be caused by the more recent practice of words in another foreign language.

Acknowledgements

We would like to thank Giulio Severijnen for help with data acquisition and preprocessing as well as Rosarita Leone for the Italian audio recordings.

References

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language*, *49*(4), 415–445. https://doi.org/10.1016/j.jml.2003.08.006

Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1063–1087. https://doi.org/10.1037/0278-7393.20.5.1063

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database [Database]. Retrieved from http://celex.mpi.nl/

Bailey, L., & Newman, A. J. (2018). Retrieval-induced forgetting and second language vocabulary acquisition: Insights from a Welsh-language training 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, NL.

Bakker, I., Takashima, A., van Hell, J. G., Janzen, G., & McQueen, J. M. (2015). Tracking lexical consolidation with ERPs: Lexical and semantic-priming effects on N400 and LPC responses to newly-learned words. *Neuropsychologia*, *79*, 33–41. https://doi.org/10.1016/j.neuropsychologia.2015.10.020

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bentin, S., & McCarthy, G. (1994). The effects of immediate stimulus repetition on reaction time and event-related potentials in tasks of different complexity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *20*(1), 130–149. https://doi.org/10.1037/0278-7393.20.1.130

Bobb, S. C., & Wodniecka, Z. (2013). Language switching in picture naming: What asymmetric switch costs (do not) tell us about inhibition in bilingual speech planning. *Journal of Cognitive Psychology, 25*(5), 568-585. https://doi.org/10.1080/20445911.2013.792822

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9/10), 341–345.

Branzi, F. M., Martin, C. D., Abutalebi, J., & Costa, A. (2014). The after-effects of bilingual language production. *Neuropsychologia*, *52*(1), 102–116. https://doi.org/10.1016/j.neuropsychologia.2013.09.022

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

Christoffels, I., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potention study. *Brain Research*, *1147*, 192–208. https://doi.org/10.1016/j.brainres.2007.01.137

Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent? *Journal of Memory and Language*, *45*(4), 721-736. https://doi.org/10.1006/jmla.2001.2793

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283. https://doi.org/10.1037/0278-7393.26.5.1283

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of memory and Language*, *50*(4), 491-511. https://doi.org/10.1016/j.jml.2004.02.002

Degani, T., Kreiner, H., Ataria, H., & Khateeb, F. (2020). The impact of brief exposure to the second language on native language production: Global or item specific? *Applied Psycholinguistics*, *41*(1), 153-183.

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

Ferreira, C. S., Marful, A., Staudigl, T., Bajo, M. T., & Hanslmayr, S. (2014). Medial prefrontal theta oscillations track the time course of interference during selective memory retrieval. *Journal of Cognitive Neuroscience*, *26*(4), 777–791. https://doi.org/10.1162/jocn_a_00523

Finnigan, S., Humphreys, M. S., Dennis, S., & Geffen, G. (2002). ERP ‘old/new’ effects: Memory strength and decisional factor(s). *Neuropsychologia*, *40*(13), 2288–2304. https://doi.org/10.1016/S0028-3932(02)00113-6

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology*, *45*(1), 152–170. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*(7), 1220-1234. Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, *38*(3), 313-338. https://doi.org/10.1006/jmla.1997.2547

Gruber, T., Malinowski, P., & Müller, M. M. (2004). Modulation of oscillatory brain activity and evoked potentials in a repetition priming task in the human EEG. *European Journal of Neuroscience*, *19*(4), 1073–1082. https://doi.org/10.1111/j.0953-816X.2004.03176.x

Hanslmayr, S., Pastötter, B., Bäuml, K.-H. T., Gruber, S., Wimber, M., & Klimesch, W. (2008). The electrophysiological dynamics of interference during the Stroop task. *Journal of Cognitive Neuroscience*, *20*(2), 215–225. https://doi.org/10.1162/jocn.2008.20020

Hanslmayr, S., Staudigl, T., Aslan, A., & Bäuml, K.-H. T. (2010). Theta oscillations predict the detrimental effects of memory retrieval. *Cognitive, Affective and Behavioral Neuroscience*, *10*(3), 329–338. https://doi.org/10.3758/CABN.10.3.329

Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, *1*(3), 213-229. https://doi.org/10.1017/S1366728998000364

Jackson, G. M., Swainson, R., Cunnington, R., & Jackson, S. R. (2001). ERP correlates of executive control during repeated language switching. *Bilingualism: Language and Cognition*, *4*(2), 169–178. https://doi.org/10.1017/S1366728901000268

Johansson, M., Aslan, A., Baüml, K.-H., Gäbel, A., & Mecklinger, A. (2007). When remembering causes forgetting: Electrophysiological correlates of retrieval-induced forgetting. *Cerebral Cortex*, *17*, 1335–1341. https://doi.org/10.1093/cercor/bhl044

Klaus, J., Lemhöfer, K., & Schriefers, H. (2018). The second language interferes with picture naming in the first language: Evidence for L2 activation during L1 production. *Language, Cognition and Neuroscience*, *33*(7), 867-877. https://doi.org/10.1080/23273798.2018.1430837

Klimesch, W., Doppelmayr, M., Russegger, H., & Pachinger, T. (1996). Theta band power in the human scalp EEG and the encoding of new information. *Neuroreport*, *7*(7), 1235–1240. https://doi.org/10.1097/00001756-199605170-00002

Kreiner, H., & Degani, T. (2015). Tip-of-the-tongue in a second language: The effects of brief first-language exposure and long-term use. *Cognition*, *137*, 106-114. https://doi.org/10.1016/j.cognition.2014.12.011

Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica*, *128*(3), 416–430. https://doi.org/10.1016/j.actpsy.2008.02.001

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, *44*, 325–343. https://doi.org/10.3758/s13428-011-0146-0

Lemhöfer, K., Huestegge, L., & Mulder, K. (2018). Another cup of TEE? The processing of second language near-cognates in first language reading. *Language, Cognition and Neuroscience*, *33*(8), 968-991. https://doi.org/10.1080/23273798.2018.1433863

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*(8), 707–710. https://doi.org/citeulike-article-id:311174

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

Mandera, P., Keuleers, E., & Brysbaert, M. (2017). Explaining human performance in psycholinguistic tasks with models of semantic similarity based on prediction and counting: A review and empirical validation. *Journal of Memory and Language*, *92*, 57–78. https://doi.org/10.1016/j.jml.2016.04.001

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of memory and language*, *40*(1), 25-40. https://doi.org/10.1006/jmla.1998.2602

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

Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language*, *67*(1), 224–237. https://doi.org/10.1016/j.jml.2012.05.001

Müller, G.E., & Pilzecker, A. (1900). Experimentelle Beiträge zur Lehre vom Gedächtnis. *Zeitschrift für Psychologie Ergänzungsband, 1*, 1-288.

Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., & Ridderinkhof, K. R. (2003). Electrophysiological correlates of anterior cingulate function in a go/no-go task: Effects of response conflict and trial type frequency. *Cognitive, Affective, & Behavioral Neuroscience, 3*(1), 17–26. https://doi.org/10.3758/CABN.3.1.17

Nigbur, R., Ivanova, G., & Stürmer, B. (2011). Theta power as a marker for cognitive interference. *Clinical Neurophysiology, 122*(11), 2185–2194. https://doi.org/10.1016/j.clinph.2011.03.030

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*, 156869. https://doi.org/10.1155/2011/156869

Piai, V., Roelofs, A., Jensen, O., Schoffelen, J.-M., & Bonnefond, M. (2014). Distinct patterns of brain activity characterise lexical activation and competition in spoken word production. *PLoS ONE*, *9*(2), e88674. https://doi.org/10.1371/journal.pone.0088674

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory & Cognition*, *18*(4), 367–379. https://doi.org/10.3758/BF03197126

Rugg, M. D., Cox, C. J. C., Doyle, M. C., & Wells, T. (1995). Event-related potentials and the recollection of low and high frequency words. *Neuropsychologia*, *33*(4), 471–484. https://doi.org/10.1016/0028-3932(94)00132-9

Rugg, M. D., & Curran, T. (2007). Event-related potentials and recognition memory. *Trends in Cognitive Sciences*, *11*(6), 251–257. https://doi.org/10.1016/j.tics.2007.04.004

Smith, M. E. (1993). Neurophysiological manifestations of recollective experience during recognition memory judgments. *Journal of Cognitive Neuroscience*, *5*(1), 1–13. https://doi.org/10.1162/jocn.1993.5.1.1

Staudigl, T., Hanslmayr, S., & Bäuml, K.-H. T. (2010). Theta Oscillations Reflect the Dynamics of Interference in Episodic Memory Retrieval. *Journal of Neuroscience*, *30*(34), 11356–11362. https://doi.org/10.1523/JNEUROSCI.0637-10.2010

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, *64*(1), 49-60.

Verhoef, K. M., Roelofs, A., & Chwilla, D. J. (2010). Electrophysiological evidence for endogenous control of attention in switching between languages in overt picture naming. *Journal of Cognitive Neuroscience*, *22*(8), 1832-1843. https://doi.org/10.1162/jocn.2009.21291

Van Assche, E., Duyck, W., & Gollan, T. H. (2013). Whole-language and item-specific control in bilingual language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1781-1792. https://doi.org/10.1037/a0032859

Wilding, E. L. (2000). In what way does the parietal ERP old/new effect index recollection? *International Journal of Psychophysiology*, *35*(1), 81–87. https://doi.org/10.1016/S0167-8760(99)00095-1

Wixted, J. T. (2004). The psychology and neuroscience of forgetting. *Annual Review of Psychology*, *55*, 235-269. https://doi.org/10.1146/annurev.psych.55.090902.141555

Wodniecka, Z., Szewczyk, J., Kałamała, P., Mandera, P., & Durlik, J. (2020). When a second language hits a native language. What ERPs (do and do not) tell us about language retrieval difficulty in bilingual language production. *Neuropsychologia*, *141*, 107390. https://doi.org/10.1016/j.neuropsychologia.2020.107390

Zheng, X., Roelofs, A., Erkan, H., & Lemhöfer, K. (2020). Dynamics of inhibitory control during bilingual speech production: An electrophysiological study. *Neuropsychologia*, *140*, 107387. https://doi.org/10.1016/j.neuropsychologia.2020.107387

Appendix

Stimulus list with Italian and English translations and condition assignments

Italian	English	Condition
albero	tree	Set 1
altalena	swing	Set 1
aquilone	kite	Set 1
ascia	axe	Set 1
cannuccia	straw	Set 1
capra	goat	Set 1
cespuglio	bush	Set 1
coltello	knife	Set 1
coperta	blanket	Set 1
cucchiaio	spoon	Set 1
cuffie	headphones	Set 1
foglia	leaf	Set 1
formica	ant	Set 1
freccia	arrow	Set 1
guinzaglio	leash	Set 1
legno	wood	Set 1
lumaca	snail	Set 1
manica	sleeve	Set 1
mattone	brick	Set 1
mosca	fly	Set 1
onda	wave	Set 1
orecchino	earring	Set 1
panchina	bench	Set 1
pipistrello	bat	Set 1
quadro	painting	Set 1
ramo	branch	Set 1
rana	frog	Set 1
scivolo	slide	Set 1
specchio	mirror	Set 1
squalo	shark	Set 1

sughero	cork	Set 1
tenda	curtain	Set 1
baffi	moustache	Set 1
bara	coffin	Set 1
canestro	basket	Set 1
fischietto	whistle	Set 2
ala	wing	Set 2
bambola	doll	Set 2
dado	dice	Set 2
accendino	lighter	Set 2
ago	nail	Set 2
arancia	orange	Set 2
cappello	hat	Set 2
cerniera	zipper	Set 2
chiave	key	Set 2
ciliegia	cherry	Set 2
cintura	belt	Set 2
ciotola	bowl	Set 2
cipolla	onion	Set 2
fiammifero	match	Set 2
frusta	whip	Set 2
gabbia	cage	Set 2
gamba	leg	Set 2
gonna	skirt	Set 2
guscio	shell	Set 2
matita	pencil	Set 2
nuvola	cloud	Set 2
pala	shovel	Set 2
pannolino	diaper	Set 2
pollice	thumb	Set 2
recinto	fence	Set 2
schiena	back	Set 2
scopa	broom	Set 2

semaforo	traffic light	Set 2
spazzola	hairbrush	Set 2
stivale	boot	Set 2
teschio	skull	Set 2
uva	grapes	Set 2
vestaglia	bathrobe	Set 2
zaino	backpack	Set 2
torta	cake	Filler
radice	root	Filler
candela	candle	Filler
sciarpa	scarf	Filler
aeroplano	airplane	Filler
arco	bow	Filler
guanto	glove	Filler
portafogli	wallet	Filler
impermeabile	raincoat	Filler
pesca	peach	Filler
sedia a rotelle	wheelchair	Filler
rasoio	razor	Filler
tacco	heel	Filler
valigia	suitcase	Filler
dente	tooth	Filler
piatto	plate	Filler
orologio	watch	Filler
bottone	button	Filler
ferro da stiro	iron	Filler
torre	tower	Filler
collana	necklace	Filler
corda	rope	Filler
cravatta	tie	Filler
sega	saw	Filler
tamburo	drum	Filler
reggiseno	bra	Filler

aglio	garlic	Filler
bottiglia	bottle	Filler
fungo	mushroom	Filler
finestra	window	Filler
francobollo	stamp	Filler
coperchio	lid	Filler
osso	bone	Filler
ponte	bridge	Filler
completo da uomo	suit	Filler

Highlights

- Foreign language (L3) attrition can be induced by the use of another language (L2).
- N2 and oscillatory theta power index this language interference at L3 test.
- The LPC, in turn, indexes its consequences: reduced L3 retrieval ease/success.
- N2 amplitude during L2 naming is directly related to later recall ability in L3.