Bilingualism: Language and Cognition

cambridge.org/bil

Language of instruction affects language interference in the third language

Brendan Tomoschuk[1], Wouter Duyck[2], Robert J. Hartsuiker[2], Victor S. Ferreira[1] and Tamar H. Gollan[1]

[1]University of California, San Diego, La Jolla, United States and [2]Ghent University, Ghent, Belgium

Research Article

Cite this article: Tomoschuk B, Duyck W, Hartsuiker RJ, Ferreira VS, Gollan TH (2021). Language of instruction affects language interference in the third language. *Bilingualism: Language and Cognition* **24**, 707–718. https://doi.org/10.1017/S1366728921000043

Received: 16 September 2019
Revised: 24 January 2021
Accepted: 27 January 2021
First published online: 17 March 2021

Keywords:
bilingualism; foreign language effect; L2 Status Factor; phoneme monitoring; language learning

Address for correspondence:
Brendan Tomoschuk,
E-mail: btomoschuk@ucsd.edu

Abstract

Applied linguistic work claims that multilinguals' non-native languages interfere with one another based on similarities in cognitive factors like proficiency or age of acquisition. Two experiments explored how trilinguals regulate control of native- and non-native-language words. Experiment 1 tested 46 Dutch–English–French trilinguals in a monitoring task. Participants decided if phonemes were present in the target language name of a picture, phonemes of non-target language translations resulted in longer response times and more false alarms compared to phonemes not present in any translation (Colomé, 2001). The second language (English) interfered more than the first (Dutch) when trilinguals monitored in their third language (French). In Experiment 2, 95 bilinguals learned an artificial language to explore the possibility that the language from which a bilingual learns a third language provides practice managing known-language interference. Language of instruction modulated results, suggesting that learning conditions may reduce interference effects previously attributed to cognitive factors.

Introduction

Choosing among multiple labels to express the same meaning is an often challenging but necessary part of knowing more than one language. When the lexicon includes two or more labels to express the same meaning, cognitive mechanisms must regulate and manage the selection of those labels automatically and quickly during speech production. Considerable research has explored the nature of these cognitive mechanisms in bilinguals, showing how co-activation of both languages interplays with inhibitory mechanisms, allowing for successful, and sufficiently fast, production of the right words in the right language (see Runnqvist, Strijkers & Costa, 2014 for review). Much less research, however, has explored how the dynamics of the system interact with the trajectory of how bilinguals learned the languages they speak, and more specifically with the mechanisms by which the new language is acquired.

Importantly, we know that both of a bilingual speaker's languages are constantly active during comprehension (e.g., Dijkstra & Van Heuven, 2002; Thierry & Wu, 2007, Van Assche, Duyck, Hartsuiker & Diependaele, 2009; see Kroll & De Groot, 2009 for review) and production (see de Bot, 2000 or Kroll & Gollan, 2013 for a review, though see Costa, Pannunzi, Deco & Pickering, 2017, for a competing account). One study revealed such dual-language competition in a production task that did not explicitly or obviously present words from both languages (Colomé, 2001). Catalan–Spanish bilinguals saw pictures and were asked to respond via key press "yes" if a specific phoneme was in the Catalan name for the picture and "no" if not. For example, when viewing a picture of a table, the bilingual should respond "yes" to the letter *t* because the Catalan name for table is *taula*. However, they should respond "no" when the letter is *m*, which is in the Spanish translation *mesa*. Response times were slower and false alarms more likely when responding to letters like *m* than *f*, as the /m/ sound is present in the bilingual's other, non-target language (Spanish, in this case) while /f/ is not present in either translation. Though this task involves monitoring phonemes cued by letters and not overt production, it suggests that even when formulating a word in Catalan, the Spanish translation of that word from the other language is active and competing for production.

If all of the languages that a bilingual knows compete in production, what is the nature of the cognitive system that regulates such competition? Green (1998) proposes that the attentional system uses cognitive control mechanisms to inhibit one language in anticipation of the production of another. This inhibitory control model suggests that bilinguals actively suppress lexical representations in the non-target language at the lemma level, thereby resolving cross-lingual interference and allowing for successful production of one language without intrusions – but how does such a mechanism develop in language learners? How does it interact with proficiency or use patterns of any given language? A potentially powerful, and relatively underutilized, way to answer these questions is to determine the patterns of interference

© The Author(s), 2021. Published by Cambridge University Press

among the three languages spoken by trilinguals (for a review of existing third language processing literature see De Bot & Jaensch, 2015). In particular, any differences that might arise between interference patterns among each of the pairs of trilinguals' three languages could reveal the principles that lead to that interference, in turn revealing how language activation is controlled more generally.

Early work exploring the nature of interference between languages in trilinguals showed that non-native languages interact more strongly with one another than with the native language, regardless of typological language similarity. In particular, the third language (L3) interferes with the second language (L2) more strongly than with the first language (L1) and vice versa, even though trilinguals are usually more proficient in their L1 than either the L2 or the L3. This phenomenon was first referred to as the "foreign language effect" in Meisel (1983), and later the "L2 Status Factor" in Williams and Hammarberg (1998). The latter researchers sought to determine the patterns of interaction among trilinguals' three languages during connected speech. They studied language production in a case study of L3 (Swedish) of an English–German–Swedish trilingual. They found that when the trilingual switched out of the L3 without any clear pragmatic purpose, they almost exclusively switched into the L2. They proposed that the L2 and L3 are activated in parallel during L3 production while the L1 is more inhibited. If two languages are activated in parallel during production, seemingly unmotivated switches between those two languages should occur more often than either would with the inhibited L1.

One possible explanation for this non-native language interaction is explored in Bardel and Falk (2012). They argued that for trilinguals, the L2 and L3 are often more "cognitively similar" (i.e., are learned in more similar circumstances) within a speaker, as opposed to more typologically similar (i.e., sharing functional and structural features) and it is this cognitive similarity that leads to transfer between L2 and L3 during early stages of L3 learning and the foreign language effect manifesting in Williams and Hammarberg (1998). They posit that the L2 and L3 often have similar ages of acquisition, learning contexts, and other environmental factors that cause the cognitive system to treat them more similarly, leading to more similar representations in the brain, causing transfer and perhaps even interference that may be attributed to non-native language status. Jiang and Forster (2001) likewise suggest that non-native languages are more likely to be stored in episodic, rather than lexical, memory and hence differ from the native language in representation and processing – an account that has received relatively little attention in the literature.

Falk and Bardel (2011) explored this cognitive similarity idea by testing French–English and English–French learners of German on object pronoun placement. In French, an object pronoun is placed before the verb (*Je* le *vois* – *I* him see), and in English it is placed after the verb (*I see* him). Interestingly, in German, object placement varies based on whether it is in a main (*Ich sehe* ihn – *I see* him) or subordinate clause (*Du weisst dass ich* ihn *sehe* – *You know that I* him *see*). Falk and Bardel found that when rating German sentences, both groups rated sentences as more acceptable when pronoun placement was similar to their respective L2, regardless of what that L2 was. They claim that because the L2 and L3 were both non-native and similar in terms of relevant cognitive factors, transfer between those two arises more than transfer between either non-native language and the L1. Puig-Mayenco, González Alonso, and Rothman (2020) systematically reviewed the literature on morphosyntactic transfer to the L3, and showed that L2 was the only source of transfer to the L3 in 60% of the studies considered. While there is evidence for many sources of transfer, a cognitive similarity theory accounts for considerable evidence. One such study, Westergaard, Mitrofanova, Mykhaylyk, and Rodina (2017), found effects of morphosyntactic transfer from the L2 Russian to the L3 English in Norwegian-Russian-English speakers' grammaticality judgements, suggesting that the similarity of the acquisition between L2 and L3 was the most influential factor when the languages were not typologically similar.

While morphosyntactic structure is typically considered to be shared between languages (Hartsuiker, Pickering & Veltkamp, 2004), there is also overlap at other processing levels including the lexical level as shown in research on cognates (e.g., *lemon* in English and *limón* in Spanish). Lemhöfer, Dijkstra and Michel (2004) presented trilinguals with Dutch–German-English cognates, which hastened lexical decision response times relative to cognates with overlap in only Dutch and German, suggesting that both non-native languages can independently impact processing in the native language (see also van Hell & Dijkstra, 2002; Costa, Caramazza & Sebastian-Galles, 2000).

Considerable literature at the lexical level has examined the role of language typology in transfer. This work shows that when learning a language that is perceived by the learner to be typologically similar to one of the previously known languages (e.g., share some phonological, morphological and/or syntactic patterns), learners will transfer knowledge from the typologically similar language rather than from distant ones (e.g., given a new verb that is a cognate with one of a bilingual's existing languages, participants will assume it shares the same syntactic properties as the known verb even when it differs, Singleton, 1987; Odlin, 1989; see Ringbom, 2007, and Ecke, 2015, for review). Trilinguals will furthermore use known lexical information to guess translations in non-native languages. For example, participants can more easily learn words with more interlingual orthographic overlap, and have greater difficulty learning false cognates (words with similar orthography but different meanings between languages) relative to words with no overlap (Vanhove & Berthele, 2015; Otwinowska & Szewczyk, 2017).

A variable that has not yet been explored in these studies, however, is the learning context in which the bilingual learned their third language (posited in Lijewska & Chmiel, 2015). Many speakers tend to learn their L3 in an L1 environment; in both Falk and Bardel (2011) and Westergaard, Mitrofanova, Mykhaylyk and Rodina (2017) data were collected in the L1 speaker environment (e.g., in Falk & Bardel, 2011, data for French–English learners were collected in a French-speaking university in Belgium while data for English–French learners were collected in an English-speaking university in Ireland). Because the typical learner has this experience, they likely have considerably more experience managing language interference between their native L1 and their non-native L3 and much less experience managing interference between their L2 and their L3. In what follows, we term the possible control benefit that might accrue between a newly learned language and the language used to learn that language an effect of LANGUAGE OF INSTRUCTION.

There has been very little work investigating whether the language of instruction used to learn an L3 indeed impacts the outcome of learning. Bogulski, Bice and Kroll (2019) investigated whether bilinguals were better at learning an L3 from one of their already known languages. They trained Spanish–English

https://doi.org/10.1017/S1366728921000043 Published online by Cambridge University Press

bilinguals, English–Spanish bilinguals, and Chinese-English bilinguals on Dutch vocabulary through English instruction. They found that bilingual learners performed better on a lexical decision task in the L3 Dutch when they had learned that L3 through their L1 English (for English–Spanish bilinguals), rather than through their L2 English (in the case of Spanish–English and Chinese–English bilinguals). Though this effect was demonstrated between language populations, and there are likely many reasons why the native Chinese-speaking group would have more difficulty learning Dutch than native English and Spanish speakers, the authors suggested that learning a new language through the L1 allows bilinguals to benefit from practice inhibiting their more dominant L1 during acquisition of L3. These results suggest that language of instruction may impact regulation of language activation.

In two experiments, we investigated foreign language and language of instruction effects in relatively low L3-proficiency Dutch–English–French trilinguals. Formal age of acquisition was similar for English and French in these participants, but proficiency in English was much higher than in French. First, in Experiment 1, we asked whether there is a foreign language effect in trilingual language interference at the lexical level. We recruited Dutch–English–French trilinguals and tested them on a lexical interference task adapted from Colomé (2001). Trilinguals did a block of phoneme monitoring in each language, determining whether or not phonemes (cued by written letters) were present in the name of the picture for the language assigned to that block. Critically, some of the to-be-monitored phonemes were selected from the translation equivalents of the target words in one of the trilinguals' other languages. For example, trilinguals saw a picture of a girl and were prompted to determine whether the /m/, /g/, or /f/ sounds (presented by the letters *m, g,* and *f* respectively) were present in the name (for /m/, yes in the L1 Dutch *meisje*, but not in the L2 English *girl,* and the L3 French *fille*; similarly for /g/ in the L2 English *girl*, and /f/ for L3 French *fille*). We expect that in general, trilinguals should be more likely to exhibit a false alarm to phonemes of a more dominant language than phonemes of a less dominant language. Dominant language representations are activated more frequently and should therefore interfere more with other languages. Additionally, if typology drives interference, we would expect to see that English and Dutch interfere more with one another because of typological similarity. Critically though, if the foreign language effect is what affects performance in the non-native language blocks, trilinguals completing the task in L2 English should more often exhibit false alarms to the (L3 French) /f/ sound than the (L1 Dutch) /m/ sound. Likewise, trilinguals should exhibit more false alarms while completing the task in the L3 French from the (L2 English) /g/ sound than from the (L1 Dutch) /m/ sound. This would suggest that the non-native languages interfere more with one another, and that the lexicon is subject to the same types of foreign language effect as other levels of language processing (see the word order effects of Bardel & Falk, 2012, discussed above).

In Experiment 2, we explored whether the foreign language effect can at least in part be explained as a language of instruction effect. Dutch–English bilinguals were trained in a novel L3 vocabulary via retrieval practice (see Rice & Tokowicz, 2020, for review of second language word learning), with the learning prompt coming either from their L1 Dutch or their L2 English. For example, in a particular trial, they might see either the word *meisje* (L1) or *girl* (L2), and were then asked to produce *karante,* the novel L3 translation that was phonologically different from both the L1 and L2, before receiving feedback (in the form of the correct answer) in the L3. After many trials like this, they performed the same monitoring task as in Experiment 1 in their novel L3. If language of instruction gives learners experience mitigating interference from the language of instruction, bilinguals monitoring the item *karante* should more often exhibit false alarms to /g/ (present in the L2 English *girl*) than /m/ (present in the L1 Dutch *meisje*) if they learned *karante* through the L1 Dutch, *meisje*. But if the new language was learned through the L2 English *girl*, phoneme interference effects should be greater between L1 Dutch and the novel L3 than between the L2 English and the novel L3. In other words, when bilinguals learn a third language, they may be improving their ability to mitigate interference from phonemes of the language of instruction more than from phonemes in the other language while monitoring in the L3. Alternatively, they may be improving their ability to mitigate all non-target language interference more generally, in which case the language of instruction should not affect interference patterns while monitoring the novel L3.

Experiment 1

Method

Participants

Dutch–English–French trilingual students (N = 46) at Ghent University, Belgium participated for credit. All trilinguals spoke Dutch natively, followed by English, learned in the classroom and reinforced via media (onset age of exposure $M = 8.63$, $SD = 3.17$), and lastly French, learned in the classroom from about the age of 6 or 7. Their average age was 18.59 (1.75) and 78% were female. Full participant characteristics, including picture naming scores (see Procedure) are shown in Table 1. Trilinguals named most pictures in their L1 Dutch, followed by L2 English, and L3 French. Though they had similar classroom education in both their non-native English and French, participants consistently demonstrated higher English proficiency than French. In Ghent, English is very common in the environment (e.g., in media, television, university courses) while French is almost only learned and used in a formal school setting.

Materials

A list of five hundred concrete nouns was used to generate the stimuli. This list was reduced to 21 items chosen on the following criteria: (1) there were no cognates among the Dutch, English, or French translations of the item, (2) the phonological forms of all of the translations started with a consonant and (3) the initial consonant phoneme of each word was not present in the other-language translations. For example, *meisje-girl-fille* satisfies these criteria because the three translations are not cognates, all words started with consonants, and the /m/ sound is not present in *girl* or *fille*, /g/ is not present in *meisje* or *fille,* and /f/ is not present in *meisje* or *girl*. In order to most strongly elicit phonological representations in the target language, letters with ambiguous grapheme to phoneme mappings were also not used (e.g., *c* in English can map to /k/ or /s/). Four items were removed after the experiment was run because they were found to have violated one of the above criteria. For each item in each language, three *yes* letters and three *no* letters were selected from the target word. The three *yes* letters were consonants in the target word. For words with insufficient unique *yes* phonemes, trials were repeated (e.g.,

https://doi.org/10.1017/S1366728921000043 Published online by Cambridge University Press

Table 1. Participant characteristics of Dutch-English-French trilinguals of Experiment 1.

	Dutch			English			French		
	M	SD	Range	M	SD	Range	M	SD	Range
% Pictures Named	81.27	14.77	34–58	43.30	18.53	7–48	10.45	6.18	2–16
Self-rated Listening	6.71	0.82	4–7	5.24	0.95	3–7	4.50	1.00	4–7
Self-rated Speaking	6.65	0.99	4–7	4.80	0.94	2–7	4.07	1.12	4–6
Self-rated Reading	6.57	1.00	3–7	5.00	0.90	2–7	4.35	1.04	3–6
Self-rated Writing	6.40	0.91	3–7	4.61	1.00	1–6	3.85	1.14	3–6

the *yes* letters for the Dutch word *jas* (coat) were *j*, *s* and *s*). Two of the *no* letters cued the initial phonemes from the other-language translations and one was a *yes* letter from another word within the same language stimuli that was not present in any of the three translations of the target word (referred to as the no-language condition). This baseline condition ensured that the frequency of different letters was relatively consistent across yes and no responses. Finally, in order to account for the variable word length and morphological complexity within each word list, each item was presented in each language and each condition to each participant. The full list of items is presented in Appendix A.

Procedure

Participants were told that they would be doing a task using all three languages that they spoke. Before being instructed on the particular task, they were given pictures of the items with Dutch, English, and French translations below the words and were asked to briefly familiarize themselves with the specific items they would be using in the experiment in all three languages (Colomé, 2001 also presented participants with the words that they were to be monitoring in Catalan).

Once ready, participants were instructed on the details of the phoneme monitoring task and given seven practice trials (with experimenter supervision). These practice trials deliberately included *yes* trials in which the letter was incorrect, but the phoneme was correct (e.g., *cow* and the letter *k* in English) and *no* trials in which the letter was correct, but the phoneme was incorrect (e.g., *shovel* and the letter *s*). While no critical trials had sounds that matched these criteria, these practice trials were used to ensure that participants knew to monitor phonemes not letters. Each trial consisted of a fixation cross appearing for 350 ms followed by a 150 ms blank screen, then a picture for 400 ms followed immediately by a letter for 600 ms. The participant had 2000 ms from the onset of the letter to respond yes or no on a button box (rightmost button was yes, leftmost no) as to whether the sound appeared in the name of the picture, using the grapheme to phoneme mapping of the target language (e.g., *w* represented the /w/ sound in the English block but represented the /ʊ/ in Dutch). After the practice, they were corrected on any mistakes and were told which language they were to use first (order fully counterbalanced between participants). Each block lasted about 6 minutes and participants could take a break for as long as they wanted before going onto the next block in the next language.

After the three language blocks were completed, trilinguals completed a picture naming task based on the Multilingual Naming Task (MINT; Gollan, Weissberger, Runnqvist, Montoya & Cera, 2012); words were chosen to represent a wide range of lexical frequencies and there were no cognates between Dutch and English. Trilinguals named the pictures in a set order and they were not prompted for alternative terms. Because the experimenter was a native speaker of English and not Dutch or French, trilinguals completed this picture naming task first in English, then Dutch, then French. This ensured that any mistakes participants made interpreting the line drawing were resolved in English. Results are shown in Tables 1 and 2. This task was used to assess language proficiency, ensuring all participants displayed similar language profiles (Tomoschuk, Ferreira & Gollan, 2018). Finally, they completed a language history questionnaire estimating their self-assessments of each language, and were debriefed on the study. This procedure was approved as part of a larger study by the UCSD Human Research Protection Program (Approval number: 140445, The Bilingual Effect on Speaking).

Analysis

Data were analyzed in R (R Core Team, 2013). All responses with a response time (RT) less than 100 ms or greater than 2000 ms (i.e., responses erroneously measured after the trial ended and before the next trial began) were removed. When analyzing response time as a dependent variable, all incorrect trials were additionally removed from the analysis.

Adjusted residual rates were calculated based on Hughes, Linck, Bowles, Koeth, and Bunting (2014). An accuracy score was calculated for each condition and participant, along with the average RT for that condition. The RT was then converted to minutes and the accuracy was divided by this value to generate an adjusted residual rate score. This metric reflects the average number of correct responses per minute per condition. For example, a participant who scored an average of 90.5% of correct responses in Dutch when the distractor was in English, and did so with an average RT of 1001 ms has an adjusted residual rate of 54.2 correct responses per minute (.905 / (1001 ms / 60,000 ms/min)) for that condition. This method helps to account for individual differences in the strategies different participants may take: in that those who respond quicker may show effects as error differences, whereas those who try to avoid errors may show effects as response time differences. This method was chosen over other methods of combining response times and accuracies (e.g., inverse efficiency scores, see Bruyer & Brysbaert, 2011) because it is considered to be robust to higher error rates, which are common in learning experiments. Though residual rates are the focus of our discussion, we first report false alarms (responding that a phoneme was present in a word when it was not) and response times as these were the a priori dependent variables.

https://doi.org/10.1017/S1366728921000043 Published online by Cambridge University Press

Table 2. Participant characteristics from Dutch-English bilinguals of Experiment 2.

	Dutch LOI						English LOI					
	Dutch			English			Dutch			English		
	M	SD	Range	M	SD	Range	M	SD	Range	M	SD	Range
% Words produced	87.68	7.54	36–62	59.7	19.46	6–51	87.21	8.51	26–54	55.96	19.09	6–52
Self-rated Listening	6.98	0.15	6–7	5.64	0.64	4–7	6.88	0.44	6–7	5.33	0.8	4–7
Self-rated Speaking	6.91	0.47	4–7	5.04	0.93	3–7	6.72	0.53	4–7	4.76	0.98	3–7
Self-rated Reading	6.93	0.25	6–7	5.51	0.79	4–7	6.93	0.25	5–7	5.64	0.74	3–7
Self-rated Writing	6.88	0.32	6–7	4.95	0.71	3–7	6.86	0.55	5–7	4.91	0.9	2–7

First, critical trials from all three language blocks were analyzed together to understand whether errors and response times differed between languages. Then within each block, models were built with Helmert contrasts (Wendorf, 2004). This allows for comparison within each language, thereby reducing noise from the relatively different proficiencies between languages. Across both experiments, there were two critical Helmert contrasts. In Contrast 1, the control condition (letters that did not appear in the non-target languages) was compared to the combination of the two critical conditions (letters appearing in the translations in the two non-target languages). In Contrast 2, phonemes from one non-target language were compared to phonemes from the other non-target language. These data were entered into a linear mixed effect model which was built with maximal random effect structures; when a model failed to converge, correlations were removed from the model (Barr, Levy, Scheepers & Tily, 2013). This method was chosen a priori in lieu of a full model (3 languages x 3 distractor conditions) to better capture the relationship between the control distractor and the two other language distractors as well as to simplify the complexity of the model to avoid complexity leading to convergence issues.

Results and discussion

Trilinguals responded correctly (to both *No* and *Yes* trials) in the Dutch language block on 84.3% ($SD = 7.02\%$) of trials, 76.9% ($SD = 10.6\%$) in the English language block, and 69.9% ($SD = 9.96\%$) of the time in the French language block.

Figure 1 shows the error rates organized by language block for critical (*no*) trials. There was a main effect of target language block such that error rates were highest in the L3 (French, $M = 18.5\%$, $SD = 11.6\%$), followed by L2 (English, $M = 14.8\%$, $SD = 12.0\%$) and L1 (Dutch, $M = 10.3\%$, $SD = 7.9\%$). This difference was significant ($\chi^2 = 10.39$, $p = .001$) and confirms our language proficiency assumptions. While monitoring in the L1, phonemes from the L2 and the L3 were significantly more likely to induce false alarms than letters from the no-language condition (Contrast 1 of the Helmert contrasts, $\chi^2 = 38.84$, $p < .001$). L2 and L3 were not differentially likely to cause false alarms (Contrast 2 of the Helmert contrast, $\chi^2 = 0.473$, $p = .492$). Likewise, in L2, L1 and L3 phonemes led to significantly more false alarms relative to the no-language condition (Contrast 1, $\chi^2 = 7.14$, $p = .008$), but not differentially (Contrast 2, $\chi^2 = 0.363$, $p = .547$). In the L3, however, L1 and L2 phonemes were, together, only marginally likely to induce false alarms relative to the no-language condition (Contrast 1, $\chi^2 = 2.91$, $p = .088$), and there was a significant difference such that L2 phonemes induced more false alarms than L1 (Contrast 2, $\chi^2 = 4.17$, $p = .041$).

Figure 2 shows response times, also organized by language block. Correct response time trials were log-transformed and analyzed by the same methods. As with the errors, response times first showed a main effect of language ($\chi^2 = 15.72$, $p < .001$), such that response times were fastest in the L1 block, slower in the L2 block, and slowest in the L3 block. Within the L1 block, response times were slower when monitoring for the phonemes present in the L2 or L3 translations relative to the no-language condition (Contrast 1, $\chi^2 = 6.79$, $p = .01$), but there was no difference when monitoring for L2 versus L3 phonemes (Contrast 2, $\chi^2 = 1.02$, $p = .312$). There were no within-language-block differences when monitoring L2 English or L3 French ($ps > .11$).

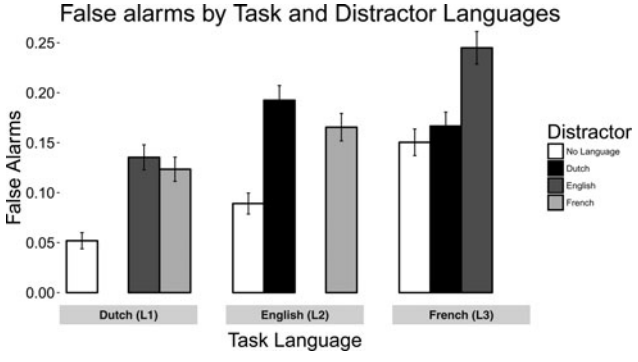


Fig. 1. False alarms grouped by whether a phoneme from the distractor language is present in the task-language name of a picture. Error bars represent standard error.

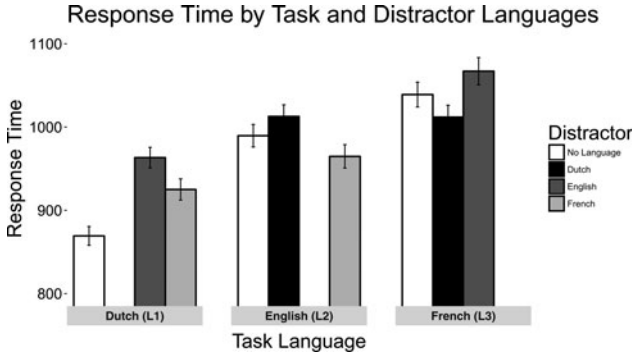


Fig. 2. Response times indicating grouped by whether a phoneme from the distractor language is present in the task-language name of a picture. Error bars represent standard error.

https://doi.org/10.1017/S1366728921000043 Published online by Cambridge University Press

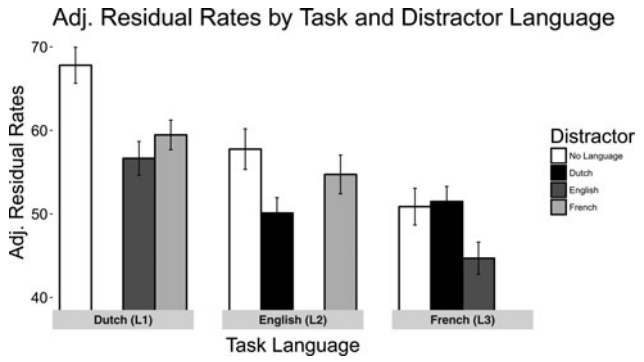


Fig. 3. Residual rates grouped by whether a phoneme from the distractor language is present in the task-language name of a picture. Higher rates represent fewer errors and faster responses. Error bars represent standard error.

To quantify any potential individual differences in strategies taken in this task, we considered speed-accuracy tradeoffs. On any given trial, a participant could decide to respond more quickly and risk making an error, or could take more time to increase the likelihood of a correct answer. We find evidence that suggests this may be occurring: in that, in a logistic regression, log response times significantly predict correct responses and interact with both language and condition ($ps < .05$). To capture these tradeoffs, we look at a combined measure of accuracy and response time: residual rates. Figure 3 shows residual rates (formula described above). Note that higher residual rates reflect overall better performance in the task (as opposed to more false alarms which indicate overall worse performance). First, there was an overall main effect of target language block ($\chi^2 = 105.96$, $p < .001$) such that rates were overall highest in the L1 block, lower in the L2 block, and lowest in the L3 block. Residual rates in the L1 were significantly worse when target phonemes appeared in the L2 and L3 translation compared to the no-language condition (Contrast 1, $\chi^2 = 69.29$, $p < .001$). Additionally, phonemes from the L2 had significantly worse residual rates relative to phonemes from the L3 (Contrast 2, $\chi^2 = 6.27$, $p = .012$). Thus, the relatively more dominant L2 interfered more than the less dominant L3 in this block. Likewise, during phoneme monitoring in the L2, there were significantly worse rates when target phonemes were from the L1 and the L3, relative to the no-language condition (Contrast 1, $\chi^2 = 17.68$, $p < .001$), and significantly worse rates when target phonemes were from the L1 than when target phonemes were from the L3 (Contrast 2, $\chi^2 = 10.33$, $p = .001$). Finally, during phoneme monitoring in the L3, residual rates were significantly worse when target phonemes were from the L1 or the L2 relative to the no-language condition (Contrast 1, $\chi^2 = 4.71$, $p = .030$), and they were significantly worse when target phonemes were from the less dominant L2 than when target phonemes were from the L1 (Contrast 2, $\chi^2 = 19.42$, $p < .001$), in contrast to the patterns of the other two language blocks. These effects did not differ significantly between participants with different block orders ($ps > .12$ for interactions between block and condition), suggesting no carry-over effects between languages.

These results demonstrate a foreign language effect in a language interference paradigm. In the residual rates (the primary focus of this analysis), we saw that, while working in the L3, target phonemes from the non-target L2 reduced rates (i.e., reduced performance) more than target phonemes from the more dominant L1. Additionally, during phoneme monitoring in the L1, target phonemes from the more dominant L2 reduced rates more than target phonemes from the less dominant L3. During phoneme monitoring in L2, target phonemes from the more dominant L1 also reduced rates more than target phonemes from the L3. While working in a higher proficiency language (L1 or L2), the more dominant language tends to interfere more, but while working in a lower proficiency L3, this pattern reverses such that the less dominant L2 tends to interfere more. This pattern suggests that, while working in a lower proficiency L3, the cognitive system engages different inhibitory mechanisms to fully suppress the L1 from when working in a higher proficiency L1 or L2. The results of this full analysis pattern identically to the aforementioned Helmert contrasts.

A similar pattern appeared in false alarms. While working in a low proficiency L3, target phonemes from the dominant L1 caused significantly fewer false alarms than target phonemes in the L2. To avoid making errors from their very dominant L1, speakers appear to have inhibited translations from their L1 especially effectively. There was not, however, an effect in response times. This may be due to the nature of the task. In the original experiments in Colomé (2001), effects were seen in response time but not (usually) in error rates. In Colomé's first experiment, the phoneme appeared for 1000 ms, followed by a blank screen for 1000 ms, and the picture for another 2000 ms. Our experiments were modeled after Colomé's third experiment (in which effects were present on response times and error rates). The phoneme appears for only 400 ms, preceded by 600 ms on the picture with no blank screen in between, which may be why the effects appeared in false alarms rather than in response times.

Interestingly, the complimentary foreign language effect that one might anticipate seeing in the L2 (such that the L3 interfered more than the L1) was not found. If the foreign language effect was really about cognitive similarity (or some other shared status between the L2 and L3), we might expect L3 to affect L2 more so than the cognitively dissimilar L1. One possible explanation for this asymmetry is consistent with a weaker foreign language effect explanation, whereby the non-native L3 is especially vulnerable to interference from the non-native L2 because these trilinguals are especially less proficient in their L3, and therefore engage in different control mechanisms to suppress the more dominant L1 (this rationale is further explored in the General Discussion). It might likewise be suggested that the complimentary foreign language effect would not appear when working in the L2 simply because the relative weakness of L3 would not lead to any interference in a stronger language. Critically, though, the L3 did still interfere when monitoring in the L2 and the L1, even in blocks where trilinguals had not yet performed the monitoring task in the L3. The L3 consistently interferes with other languages at similar levels as the other competing language – suggesting that while the L3 is relatively weak in these trilinguals, it is strong enough to consistently interfere in other languages during this task.

Additionally, we found no evidence of an impact of language typology. Though there were no cognates or homographs in the stimuli, one might expect that participants' previous experience with these languages induced interference in more typologically similar languages (i.e., Dutch and English). However, in Experiment 1, there was more interference from L2 English than L1 Dutch while producing the L3 French, and the L3 French elicited interference in both the L1 Dutch and the L2 English. If previous knowledge of language typology were the

https://doi.org/10.1017/S1366728921000043 Published online by Cambridge University Press

sole factor causing interference, English and Dutch would have interfered with one another and neither would have impacted performance in French, and French should not have impacted English or Dutch.

A different explanation that can more fully account for these data is that these trilinguals learned their L3 in an L1 classroom and surrounding environment – a language-of-instruction effect, as described in the introduction. Because trilinguals were not taught L3 through their L2, they have relatively little experience inhibiting L2 while working in L3. Indeed, foreign language effects in general could be explained by the participant's usage of their two known languages while acquiring the third. Real-life language acquisition often occurs through L1. Therefore, in order to examine this explanation, in Experiment 2, we experimentally manipulate language of instruction during acquisition of an artificial language and observe how it impacts lexical regulation in the same phoneme monitoring task.

Experiment 2

In Experiment 1, we saw disproportionate interference from the L2 while monitoring phonemes in the L3. This pattern may support a foreign language effect explanation, whereby L3 suffers more interference from L2 than L1, because L3 and L2 are more similar in cognitive profile. Alternatively, it may be due to the fact that these trilinguals' L3 was learned through L1, allowing them to better learn to inhibit L1 than L2 when speaking and monitoring in L3. To explore this language-of-instruction explanation, Dutch–English bilinguals learned new L3 items, either via their L1 or L2. If language of instruction impacts lexical interference, the language of instruction should lead to less interference than the alternative language. In other words, bilinguals who learn L3 via L1 should show more interference in L3 from L2 (as in Experiment 1), but bilinguals who learned L3 via L2 should show more interference in L3 from L1. This latter pattern would be the reverse of what should be observed due to a foreign language effect explanation. If language of instruction cannot explain this effect, and instead the disproportionate interference in Experiment 1 is the result of similarity in the mutual cognitive profile of the non-native languages, L2 should interfere with L3 more regardless of the language of instruction. Incidentally, an alternative, simple associative account might predict the opposite pattern: if the L3 is learned through the L1, then the L3 and L1 words could become associated and so could activate each other. This predicts that, if L3 is learned via L1, L3 monitoring should be more difficult for an L1 phoneme – the opposite of the language of instruction account (which predicts that, in this case, monitoring should be more difficult for an L2 phoneme).

Method

Participants

Dutch–English bilinguals (N = 95) recruited from Ghent University participated for course credit or payment. All bilinguals were Dutch dominant. Participants were not recruited or tested based on the knowledge (or lack thereof) of a third language and are therefore referred to as bilinguals. Five participants were removed from the analysis for incorrect performance on the task, and a further sixteen were removed for not being able to surpass chance performance, with 60% errors on the phoneme monitoring task (the significance threshold in a binomial test of the same number of trials), leaving 74 participants. This rate of people performing at or below chance is likely due to the difficult nature of the combination of the learning and monitoring tasks. Further participants were therefore added until there were 95 that performed above chance, 46 who learned via Dutch, and 49 who learned via English.[1] Participants in the final Dutch sample were 91.1% female, an average age of 18.8 (1.8) and were first exposed to English at 8.2 (4.0). Bilinguals in the final English sample were 75.5% female, an average of 20.1 (5.8) years old and first exposed to English at 8.5 (4.0). Full participant characteristics are shown in Table 2.

Procedure

Bilinguals were told that they would do a task that involved 20 Dutch and English words, and they were shown the Dutch and English names, and given as much time as they requested to study them before continuing. Participants were then told they would learn these words in a new language, called Ibararpa. Artificial words were taken by using Italian pseudowords generated in Wuggy, a pseudoword generator, to ensure naturalistic items (Keuleers & Brysbaert, 2010) that were about equally similar to Dutch and English. Italian was chosen because it is phonetically transparent and not typologically similar to either known language. They were not told that the language was artificial until the end of the experiment. In each of four learning blocks, they had a brief exposure to each word in Ibararpa, by viewing the picture of the item from the familiarization phase and hearing Ibararpa audio of the word. They had as much time as they required to reach familiarity on each word before continuing. After being exposed to each word once, they began the training phase. In this phase, half of the participants saw a Dutch word appear on the left side of the screen, and heard the Dutch audio of the word and had four seconds to speak the Ibararpa word. After the four second delay period, the Ibararpa word appeared on the right, also accompanied by Ibararpa audio of the word.

Based on pilot data collected to determine the optimal learning structure, a between-subject design in which some participants learned via Dutch and some learned via English was chosen for this study. Bilinguals learned words in groups of 5, that they practiced retrieving 8 times in a block. Between blocks, they had a break that could last as long as they wanted. At the end, they had one final block that tested their knowledge. They saw each word they had learned presented in the same method (with Dutch prompts and Ibararpa feedback). They only saw each word once. The other half of the participants (determined by random assignment) completed this same task but used English-based prompts in learning rather than Dutch. This variable was manipulated between participants for two reasons: first, a within-subject experiment in which a participant learned some translations from Dutch and others from English would not accurately represent how the trilinguals in Experiment 1 had learned French. Second, a within-subject design would require participants to learn twice as many translations and therefore, based on pilot data, require considerably more training time.

At the end of the training, participants performed one block of phoneme monitoring in Ibararpa. The procedure for this task was identical to the monitoring blocks of Experiment 1. After this, they completed the MINT in Dutch and English and completed a shortened Language History Questionnaire.

[1]Note that applying the same filtering procedure to Experiment 1 does not alter the outcomes in any way.

https://doi.org/10.1017/S1366728921000043 Published online by Cambridge University Press

Analysis

The analysis was similar to Experiment 1. Helmert contrasts were used to first assess whether there was any effect due to phonemes from the no-language condition, then to assess whether there were differential effects of the two phoneme languages. Language of instruction was also added as a factor. Additionally, Wald Z tests were used to assess model significance.

Results and discussion

Words were considered to be learned if in the final instructional block they were produced within one phoneme of the target word (so an utterance that is only one phoneme different from the target, like *karanta,* was considered a correct production of the target *karante* but an utterance that differed by two or more phonemes like *kamanta* was considered incorrect). Overall, bilinguals who learned via the L1 Dutch scored slightly higher on the final block of learning ($M = 63.6\%$, $SD = 14.9\%$) than those who learned via the L2 English ($M = 62.8\%$, $SD = 16.2\%$), but this difference was not statistically significant ($t = 0.25$, $p = .80$). Additionally, those who learned via Dutch made slightly fewer errors in the monitoring task ($M = 25.9\%$, $SD = 7.62\%$) across all trials than those who learned via English ($M = 27.8\%$, $SD = 7.54\%$), though again the difference was not significant ($t = 1.24$, $p = .221$). The response time difference across all trials between the Dutch learners ($M = 945$, $SD = 168$) and the English learners ($M = 969$, $SD = 170$) was not significant ($t = -0.70$, $p = .484$).

Figure 4 shows false alarms for Experiment 2. Because the monitoring task was conducted in only the artificial L3, the x-axes in these graphs show the language from which bilinguals learned the new language. Here, we see a significant main effect of interfering language, such that phonemes from Dutch and English induced more errors than phonemes from the no-language condition (Contrast 1, $z = -2.20$, $p = .028$), across language of instruction. Bilinguals who learned through Dutch made slightly more errors with target phonemes from English while bilinguals who learned via English did not seem to have more errors with the target phoneme from Dutch or English, though this interaction between language of instruction and Contrast 2 was not significant ($z = 0.93$, $p = .35$). When analyzing each language of instruction group individually (i.e., when analyzing the data from just the Dutch or English language of instruction groups), there was a significant effect in the error rates when the language of instruction was English, such that phonemes from Dutch and English were more likely to false alarm than those from neither translation (Contrast 1, $z = -2.22$, $p = 0.026$). In the response times (Figure 5), there was a marginal effect when the language of instruction was Dutch, such that phonemes from the L2 English were responded to slower than phonemes from the L1 Dutch (Contrast 2, $t = 1.88$, $p = .077$). There were no significant effects when looking at the error rates of the Dutch language-of-instruction group or the response times of the English language-of-instruction group ($ps > 0.27$).

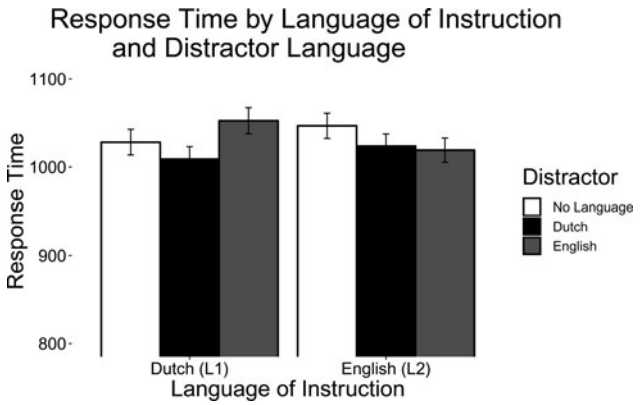


Fig. 5. Response times grouped by whether a phoneme from the distractor language is present in the artificial language, and by language of instruction. Error bars represent standard error.

Figure 6 shows residual rates for the monitoring task of Experiment 2. As in Experiment 1, we analyzed speed-accuracy tradeoffs and found that logged response times were a significant predictor of correct responses, and that they significantly interact with language of instruction ($p < .001$). As such, we analyzed residual rates. There was a marginal effect, such that rates were lower (i.e., performance was worse) with target phonemes from Dutch and English compared to target phonemes from no language (Contrast 1, $t = -1.83$, $p = .069$). There was an effect in the second contrast such that rates were higher with target phonemes from Dutch, relative to target phonemes from English (Contrast 2, $t = -4.01$, $p < .001$). Finally, there was a significant interaction between language of instruction and the second Helmert contrast, such that the rate difference (i.e., higher rates

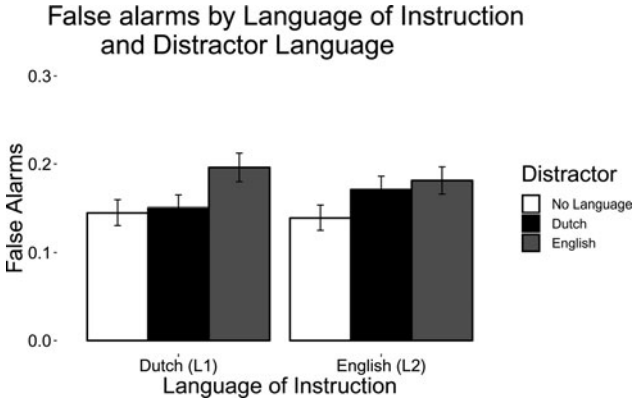


Fig. 4. False alarms grouped by whether a phoneme from the distractor language is present in the artificial language, and by language of instruction. Error bars represent standard error.

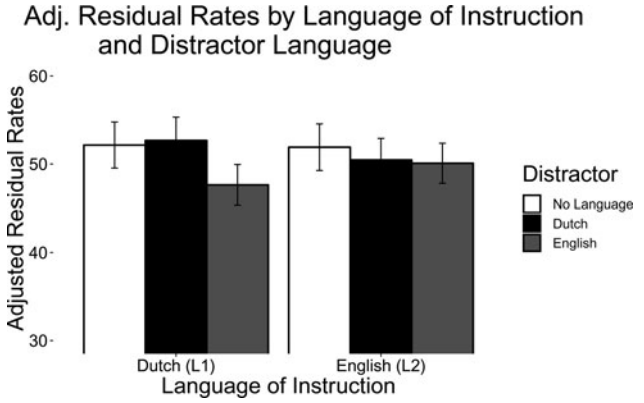


Fig. 6. Residual rates grouped by whether a phoneme from the distractor language is present in the artificial language, and by language of instruction. Error bars represent standard error.

https://doi.org/10.1017/S1366728921000043 Published online by Cambridge University Press

with target phonemes from Dutch than English) was greater for bilinguals who learned the new language from Dutch relative to those who learned from English. The latter did not show differential patterns based on which language the target phoneme was from (Contrast 2, $t = 2.67$, $p = .008$). This difference also appeared when analyzing each language of instruction group individually. When learning occurred via Dutch, the first Helmert contrast was marginally significant ($t = -1.74$, $p = .085$) and the second contrast was significant ($t = -3.82$, $p < .001$). When learning occurred via English, there were no effects in either contrast ($t = -1.63$, $p = .107$ for the first contrast and $t = 0.322$, $p = .748$ for the second).

In these results, we expected that while learning, participants would learn to inhibit the language they learned from, leaving the other language able to interfere while monitoring. While we found this pattern when they learned via their L1 (their L2 interfered in monitoring the new L3), we did not see the reverse pattern when bilinguals learned from their L2, such that their L1 interfered more in L3 production than L2. Instead, for these learners, there is virtually no interference when working in the L3 and monitoring the L1 or L2. This pattern of data suggests that language of instruction does affect interference patterns in phoneme monitoring, which may explain some effects previously attributed to the foreign language effect.

General discussion

Two experiments demonstrate that foreign language effects can be shown in language interference tasks and, second, that this effect may be explained in part by language of instruction. In Experiment 1, Dutch–English–French trilinguals performed a phoneme monitoring task, in which they monitored for specific target phonemes in all of their target languages (Colomé, 2001). We observed that phoneme monitoring in L3 (e.g., picture *fille*) was worse when pictures contained target phonemes (e.g., /g/) present in their irrelevant L2 translations (e.g., *girl)*, than when target phonemes appeared in their more proficient L1 (e.g., /m/ and *meisje*). In Experiment 2, we tested the possibility that such an effect might actually be driven by a learner's language of instruction. Bilinguals learned a novel L3 through either their L1 or L2. Afterwards, in the phoneme monitoring task, we again observed that bilinguals who learned through L1 suffered more interference from L2 than from L1 in the L3 phoneme monitoring task. Most interestingly, this effect was not found when bilinguals learned L3 through L2.

The interaction between the L1/L2 phonemes and language of instruction indicates that language of instruction does impact language interference, at least for languages at low proficiency levels. Though we tested almost 100 bilinguals in an effortful and time-intensive language learning experiment, this difference was only significant in the residual rate data, with marginal effects in the false alarms and response times when each language was considered individually. This may be due to a speed-accuracy tradeoff that varies across individuals, condition to condition, or trial to trial. Some individuals respond faster and show effects in errors, while others slow overall response time and make fewer errors, and therefore show effects in response time. This would perhaps not be surprising given the outcomes of Colomé (2001). Across the Colomé (2001) experiments, the interference effect was found in response times. In the third experiment, in which the time spent on each trial was shortened, however, the effect was also found in false alarms. The paradigm in the experiments reported here mimics the Colomé (2001) experiment in which the significant differences were found in both response times and false alarms, and so it is possible that different participants take different strategies on different trials, leading to results being most clear when both dependent variables are considered jointly.

An additional difference worth noting between this work and that of Colomé (2001) is that the languages used in the original task, Spanish and Catalan, have shallow orthographies, while Dutch, English, and French have more ambiguous mappings between orthography and phonology. A core assumption of this phoneme monitoring task is that letters presented on the screen elicit phonological activation. While materials were chosen to minimize this ambiguity (e.g., avoiding *c* in English as it maps to either the /k/ or /s/ sound), the languages of the trilinguals studied here could reduce the amount of phonological activation, and therefore interference, induced by the letters presented in this task relative to what might be seen in a study with Catalan–Spanish bilinguals.

That non-native languages share cognitive resources based on cognitive similarity is a sensible explanation to a commonly reported effect in learners of a third (or later) language. Non-native languages, especially those learned as an adult, have a particular cognitive similarity that is distinct from the native language, and it is reasonable to theorize that this makes them more likely to interfere with each other when acquiring items in a new language. Non-native languages are often acquired at similar ages, they are less dominant than the native language, and they are similar in other cognitive factors known to impact language interference. There is, however, an alternative explanation that can also account for foreign language effects (in addition to or instead of cognitive similarity): namely, language of instruction. By testing language learners who all learned their L2 and L3 in an L1 environment, previous studies could not consider language of instruction in the language learning process as a possibly critical factor. Thus, multilinguals tested in those studies had as much experience inhibiting L1 information while working in their non-native language as they had experience in that non-native language. The results of Experiment 2 suggest that this practice, and not exclusively non-native language status per se, can have significant consequences for interference in production of a new language. This language of instruction effect is likely at least a partial explanation of foreign language effects seen in other published studies (e.g., Williams & Hammarberg, 1998).

Though we experimentally manipulated language of instruction, this effect could instead represent something more like language of the general environment. The trilinguals in Experiment 1 did more than just learn their L3 in an L1 classroom, they were living, working, and studying other topics in their dominant L1. We know that the activation of one of a bilingual's two languages can be boosted, and that this affects their language processing more generally. Elston-Güttler, Gunter, and Kotz (2005) showed German-English bilinguals a twenty-minute-long video subtitled in either German or English before performing a semantic priming etask entirely in their L2 English. Participants performed a lexical decision task after reading a sentence with an interlingual homograph (e.g., *gift* is German for poison). The authors found in both behavioral and neurocognitive measures that semantic priming effects in the first block of the experiment were mediated by the language in which the video was subtitled, despite the identical test materials. This suggests that global language activation can be altered based on a more local environment. Indeed, in Experiment 2 we had bilinguals working in one of their two

https://doi.org/10.1017/S1366728921000043 Published online by Cambridge University Press

languages throughout the majority of the experiment. Our effect, then, may in part be due to global language activation while learning an L3, rather than or in addition to specific regulation between translation pairs of two languages. Put another way, the language of the learning environment may allow bilinguals to rely on their dominant L1 even when in the L2 learning condition. Further studies on language of environment may also help us understand the impact of immersion on language interference and clarify the results of Experiment 2.

Language of environment (or generally language dominance) may also explain why bilinguals who learned via their L2 did not show L1 interference in Experiment 2. While we manipulated language of instruction, we were not able to manipulate the entire language environment: all bilinguals were immersed, and dominant, in their L1, likely making it more important to inhibit their very active L1, even when they did not learn from those translation pairs. These data suggest that multilinguals are better able to inhibit their dominant language while monitoring in a less proficient non-native language, possibly especially when immersed in that dominant language. This, in combination with a language of instruction effect, would explain the pattern of results seen in Experiment 2. Thus, future work can explore this limitation by manipulating the environment and language dominance of participants to understand the impact on language of instruction, and more generally, language interference in early stages of learning. Future work might also develop a paradigm in which language of instruction is manipulated within individuals to reduce the impact of variables like language of environment or an individual's language dominance.

The results of Experiment 1 also suggest that the low proficiency L3 is more susceptible to interference from other languages. This idea was explored in Bartolotti and Marian (2019). Spanish–English participants were taught vocabulary in an artificial language that conflicted with either English or Spanish letter-sound mappings. Accuracy improved in this task over time from both languages, suggesting that while English and Spanish orthography can cause interference in the third language, learners develop mechanisms to control this interference over time. In our results, participants in a relatively low proficiency (Experiment 1) or new (Experiment 2) L3 show a considerable amount of interference from their non-target languages. If these interference patterns were to change with increasing proficiency, as Bartolotti and Marian (2019) show, this would suggest that learners develop control mechanisms with proficiency to allow speakers to produce the target language without interference from the native language, and, based on Experiment 2, that this early L3 interference can be impacted by the type, and not just the amount, of experience that a learner receives. More specifically, the asymmetry of the effects seen in Experiment 2 (i.e., that those who learned from Dutch showed interference from English but those that learned from English did not show interference from either English or Dutch) suggests that the inhibitory control mechanism can be applied to a multilingual's other languages when the L3 is taught via that other language. However, because the L1 did not interfere at all, learners must not need to specifically train inhibition of their dominant L1 while learning. We speculate that the L1 must be dominant enough to warrant inhibition in a low proficiency language regardless of language of instruction. Because phoneme monitoring is not explicitly a language control task, however, these mechanisms would need to be tested in additional paradigms to get a better understanding of the control mechanisms involved in mitigating interference.

If, though, low proficiency languages engage a different control mechanism from high proficiency ones, what is the nature of this mechanism? We speculate that the results shown here suggest that multilinguals need to use strong top-down control mechanisms to prevent dominant languages from interfering in a low proficiency language task. One could argue, based on the simple associative account (see Introduction), that the amount of interference seen in this task should be greater from the language of instruction if the connections are simple associations. The experiments here, however, suggest that speakers use top-down control mechanisms to inhibit the most intrusive language, and that language of instruction helps provide experience facilitating that inhibition.

These studies support the notion that learning lexical items in a new language involves more than just connecting a new word to a known concept; it first involves using learning experience to regulate the connection between those words to allow for successful production in a low proficiency language, and eventually a more direct connection between the new word and the conceptual representation. If a speaker of a second, third, or any new language hopes to reach a level of proficiency in which they do not rely on their language of instruction as a translational scaffold, they must down-regulate the connection between the two lexical items to allow for a stronger link to conceptual representations. Our results suggest that language control mechanisms may differ at earlier stages of acquisition. They also demonstrate that the foreign language effect can be partially explained by this language of instruction, and that this learning experience is critical in first establishing that link between a known word and a new translation.

Acknowledgements. The authors thank Merel Muylle for help recording stimuli. This research was supported by grants from the National Institute on Deafness and Other Communication Disorders (011492), the National Institute of Child Health and Human Development (050287, 051030, 079426), the National Science Foundation (BCS1457159) and a concerted research action, Grant No. BOF13/GOA/032 from Ghent University.

References

Bardel C and Falk Y (2012) The L2 Status Factor and the Declarative/Procedural distinction. *Third language acquisition in adulthood* **46**, 61–78. https://doi.org/10.1075/sibil.46.02ama

Barr DJ, Levy R, Scheepers C and Tily HJ (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

Bartolotti J and Marian V (2019) Learning and processing of orthography-to-phonology mappings in a third language. *International Journal of Multilingualism* **16**(4), 377–397. https://doi.org/10.1080/14790718.2017.1423073

Bogulski CA, Bice K and Kroll JF (2019) Bilingualism as a desirable difficulty: Advantages in word learning depend on regulation of the dominant language. *Bilingualism: Language and Cognition* **22**(5), 1052–1067. https://doi.org/10.1017/s1366728918000858

Bruyer R and Brysbaert M (2011) Combining speed and accuracy in cognitive psychology: is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)?. *Psychologica Belgica* **51**(1), 5–13. https://doi.org/10.5334/pb-51-1-5

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 and 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

https://doi.org/10.1017/S1366728921000043 Published online by Cambridge University Press

Costa A, Pannunzi M, Deco G and Pickering MJ (2017) Do bilinguals automatically activate their native language when they are not using it?. *Cognitive science* **41**(6), 1629–1644. https://doi.org/10.1111/cogs.12434

De Bot K (2000) A bilingual production model: Levelt's "speaking" model adapted. *The bilingualism reader* 420–442. https://doi.org/10.4324/9781003060406-37

De Bot K and Jaensch C (2015) What is special about L3 processing?. *Bilingualism: Language and Cognition* **18**(2), 130–144. https://doi.org/10.1017/s1366728913000448

Dijkstra T and Van Heuven WJ (2002) The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and cognition* **5**(3), 175–197. https://doi.org/10.1017/s1366728902003012

Ecke P (2015) Parasitic vocabulary acquisition, cross-linguistic influence, and lexical retrieval in multilinguals. *Bilingualism: language and cognition* **18**(2), 145–162. https://doi.org/10.1017/s1366728913000722

Elston-Güttler KE, Gunter TC and Kotz SA (2005) Zooming into L2: Global language context and adjustment affect processing of interlingual homographs in sentences. *Cognitive Brain Research* **25**(1), 57–70. https://doi.org/10.1016/j.cogbrainres.2005.04.007

Falk Y and Bardel C (2011) Object pronouns in German L3 syntax: Evidence for the L2 status factor. *Second Language Research* **27**(1), 59–82. https://doi.org/10.1177/0267658310386647

Gollan TH, Weissberger GH, Runnqvist E, Montoya RI and Cera CM (2012) Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition* **15**(3), 594–615. https://doi.org/10.1017/s1366728911000332

Green DW (1998) Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and cognition* **1**(2), 67–81. https://doi.org/10.1017/s1366728998000133

Hartsuiker RJ, Pickering MJ and Veltkamp E (2004) Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological science* **15**(6), 409–414. https://doi.org/10.1111/j.0956-7976.2004.00693.x

Hughes MM, Linck JA, Bowles AR, Koeth JT and Bunting MF (2014) Alternatives to switch-cost scoring in the task-switching paradigm: Their reliability and increased validity. *Behavior research methods* **46**(3), 702–721. https://doi.org/10.3758/s13428-013-0411-5

Jiang N and Forster KI (2001) Cross-language priming asymmetries in lexical decision and episodic recognition. *Journal of Memory and Language* **44**(1), 32–51. https://doi.org/10.1006/jmla.2000.2737

Keuleers E and Brysbaert M (2010) Wuggy: A multilingual pseudoword generator. *Behavior research methods* **42**(3), 627–633. https://doi.org/10.3758/brm.42.3.627

Kroll JF and De Groot AM (Eds.) (2009) *Handbook of bilingualism: Psycholinguistic approaches*. Oxford University Press. https://doi.org/10.1353/lan.2008.0062

Kroll JF and Gollan TH (2013) Speech planning in two languages. In *The Oxford Handbook of Language Production*. https://doi.org/10.1093/oxfordhb/9780199735471.013.001

Lemhöfer K, Dijkstra T and Michel M (2004) Three languages, one ECHO: Cognate effects in trilingual word recognition. *Language and cognitive processes* **19**(5), 585–611. https://doi.org/10.1080/01690960444000007

Lijewska A and Chmiel A (2015) Cognate facilitation in sentence context–translation production by interpreting trainees and non-interpreting trilinguals. *International Journal of Multilingualism* **12**(3), 358–375. https://doi.org/10.1080/14790718.2014.959961

Meisel JM (1983) Transfer as a second-language strategy. *Language & communication* **3**(1), 11–46. https://doi.org/10.1016/0271-5309(83)90018-6

Odlin T (1989) *Language transfer: Cross-linguistic influence in language learning*. Cambridge University Press. https://doi.org/10.1177/026765839100700305

Otwinowska A and Szewczyk JM (2017) The more similar the better? Factors in learning cognates, false cognates and non-cognate words. *International Journal of Bilingual Education and Bilingualism*. https://doi.org/10.1080/13670050.2017.1325834

Puig-Mayenco E, González Alonso J and Rothman J (2020) A systematic review of transfer studies in third language acquisition. *Second Language Research* **36**(1), 31–64. https://doi.org/10.1177/0267658318809147

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rice CA and Tokowicz N (2020) A Review of Laboratory Studies of Adult Second Language Vocabulary Training. *Studies in Second Language Acquisition* **42**(2), 439–470. https://doi.org/10.1017/s0272263119000500

Ringbom H (2007) *Cross-linguistic similarity in foreign language learning* (Vol. 21). Clevedon: Multilingual Matters.

Runnqvist E, Strijkers K and Costa A (2014) Bilingual Word Access. *The Oxford handbook of language production* **182**. https://doi.org/10.1093/oxfordhb/9780199735471.013.028

Singleton D (1987) Mother and other tongue influence on learner French: A case study. *Studies in Second Language Acquisition* **9**(3), 327–345. https://doi.org/10.1017/s0272263100006719

Thierry G and Wu YJ (2007) Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences* **104**(30), 12530–12535. https://doi.org/10.1073/pnas.0609927104

Tomoschuk B, Ferreira VS and Gollan TH (2018) When a seven is not a seven: Self-ratings of bilingual language proficiency differ between and within language populations. *Bilingualism: Language and Cognition* 1–21. https://doi.org/10.1017/s1366728918000421

Van Assche E, Duyck W, Hartsuiker R and Diependaele K (2009) Does bilingualism change native-language reading? Cognate effects in sentence context. *Psychological Science* **20**(8), 923–927. https://doi.org/10.1111/j.1467-9280.2009.02389.x

Van Hell JG and Dijkstra T (2002) Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic bulletin & review* **9**(4), 780–789. https://doi.org/10.3758/bf03196335

Vanhove J and Berthele R (2015) The lifespan development of cognate guessing skills in an unknown related language. *International Review of Applied Linguistics in Language Teaching* **53**(1), 1–38. https://doi.org/10.1515/iral-2015-0001

Wendorf CA (2004) Primer on multiple regression coding: Common forms and the additional case of repeated contrasts. *Understanding Statistics* **3**(1), 47–57. https://doi.org/10.1207/s15328031us0301_3

Westergaard M, Mitrofanova N, Mykhaylyk R and Rodina Y (2017) Crosslinguistic influence in the acquisition of a third language: The Linguistic Proximity Model. *International Journal of Bilingualism* **21**(6), 666–682. https://doi.org/10.1177/1367006916648859

Williams S and Hammarberg B (1998) Language switches in L3 production: Implications for a polyglot speaking model. *Applied linguistics* **19**(3), 295–333. https://doi.org/10.1093/applin/19.3.295

https://doi.org/10.1017/S1366728921000043 Published online by Cambridge University Press

Appendix A. Items used in Experiments 1 and 2.

Experiment 1			Experiment2		
English	Dutch	French	English	Dutch	Ibararpa
girl	meisje	fille	girl	meisje	karante
lighthouse	vuurtoren	phare	lighthouse	vuurtoren	fosi
watch	horloge	montre	leg	been	rapo
horse	paard	cheval	horse	paard	borante
smoke	rook	fumée	smoke	rook	bimo
cheese	kaas	fromage	cheese	kaas	gasmirgo
backpack	rugzak	sac à dos	backpack	rugzak	wimu
coat	jas	manteau	coat	jas	pilo
knight	ridder	chevalier	rock	steen	maipa
butterfly	vlinder	papillon	butterfly	vlinder	gotote
feather	veer	plume	feather	veer	kome
fox	vos	renard	pig	varken	siago
trashcan	vuilbak	poubelle	trashcan	vuilbak	zossigi
binoculars	verrekijker	jumelles	binoculars	verrekijker	dototemi
drill	boormachine	percer	drill	boormachine	zoika
deer	hert	cerf	deer	hert	folze
pitcher	kan	lanceur	fox	vos	woddi
			window	raam	tispibi
			tree	boom	poliri
			carrot	wortel	villibu

https://doi.org/10.1017/S1366728921000043 Published online by Cambridge University Press