Stimulus and Response Conflict from a Second Language:

Stroop Interference in Weakly-Bilingual and Recently-Trained Languages

Author notes

R scripts and data for the reported analyses are available on the Open Science Framework (link: https://osf.io/ch8x6/).

Abstract

The aim of the present manuscript was to investigate the source of congruency effects in weak bilinguals (Experiment 1) and in early language learning (Experiment 2). In both studies, participants performed a bilingual version of a colour-word Stroop task. The standard finding is slower and less accurate responding when the word and colour are incongruent (e.g., “red” in blue) relative to congruent (e.g., “red” in red). This congruency effect occurs for the distracting colour words from both the first and second language. Both stimulus conflict (i.e., conflict between the meaning of the word and ink colour) and response conflict (i.e., conflict between possible response options) contribute to first-language congruency effects. According to some models of early language learning, only one of these two types of conflict should emerge for non-fluent languages. To separate stimulus and response conflict, we used a 2-to-1 keypress assignment manipulation. Interestingly, in one study both stimulus and response conflict were evidenced for the weakly spoken second language (English in native French speakers). In a second study, participants performed a short Croatian colour word learning phase before the Stroop procedure. Stimulus conflict was observed in response times and response conflict in errors for this recently-trained language. These findings suggest that the relatively low-proficient second language words are potent enough to affect semantic identification and response selection.

Keywords: weak bilinguals, Stroop effect, stimulus conflict, response conflict, novel word acquisition, language learning

1. Introduction

Within the literature on bilingual cognition (Bialystok et al., 2012; Grosjean, 2012; Valian, 2015), a critical issue is how the second language (L2) influences cognitive processing in bilinguals, and whether this is similar or different than the first language (L1). The possible similarities or differences in the processing of L1 and L2 words have been observed in terms of their impact on semantic identification and response decision processes. There are some findings that emphasize the similarities in semantic and response processing for both first and second languages (Schmidt et al., 2018). However, this can vary as a function of L2 proficiency. The present study further investigates these similarities in processing of L1 and weakly spoken L2 words with a bilingual version of the colour-word Stroop task.

In a classical Stroop task, participants are instructed to respond to the print colour of colour words (e.g., the word "red" printed in blue ink, to which the participant should respond by saying "blue"). The basic finding, termed a *congruency* or *Stroop* effect, is that participants respond faster and more accurately when the colour word and ink colour are *congruent* (e.g., the word "red" printed in red), in contrast to *incongruent* trials in which the word and colour mismatch (e.g., the word "red" printed in blue). Much research has focused on the source of this conflict in the cognitive system. *Stimulus conflict* is a conflict in meaning between the word and the colour (e.g., that "red" and "blue" refer to different colour concepts; Glaser & Glaser, 1989). *Response conflict* is a conflict between the response elicited by the word and the response elicited by the colour (e.g., that the word "red" should be pronounced as "red" and the colour blue should be named as "blue"). Although an early debate centered on whether congruency effects were due to one or the other type of conflict (MacLeod, 1991; Stroop, 1935), the general consensus is that both stimulus and response conflict contribute to the Stroop effect (Augustinova et al., 2019; De Houwer, 2003; Klein, 1964; Schmidt &

Cheesman, 2005; Sharma & McKenna, 1998). There are other sources of conflict, such as *task conflict*, which arises from drawing attention to an irrelevant task (i.e., word reading) instead to a relevant task (i.e., colour naming), causing competition between two possible responses. That is, readable stimuli (e.g., “dog”) produce larger interference that non-readable (e.g., “xxxx”) ones (Augustinova et al., 2019; Kalanthroff et al., 2013; Monsell et al., 2001), though this is less relevant for the current series of experiments where all distracting stimuli are readable. Taken together, there is a general consensus that reaffirms a multicomponent nature of the Stroop interference effect (Augustinova et al., 2019; Klein, 1964; Neely & Kahan, 2001). The present manuscript will focus on stimulus and response conflict exclusively.

One of the clearest lines of evidence for both stimulus and response conflict contributing to the Stroop effect comes from the 2-to-1 mapping procedure. In this variation of the traditional Stroop task, two colours are mapped to each response button (De Houwer, 2003, 2004). For instance, participants might be asked to respond with the left key to blue or green targets, and the right key for red or yellow targets (Hasshim & Parris, 2014; Shichel & Tzelgov, 2018), as illustrated in Figure 1. The aim of this manipulation is to distinguish stimulus and response conflict.

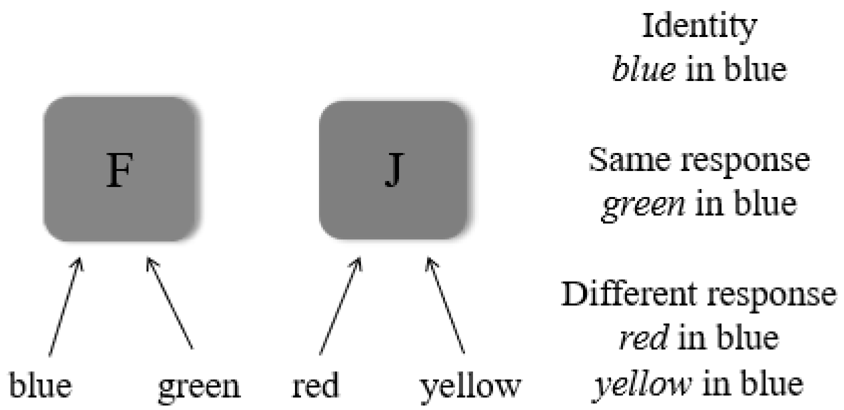

Fig. 1. Illustration of the 2-to-1 mapping procedure.

By applying the 2-to-1 mapping procedure, three possible trial types occur. First, there are *identity trials*, in which the word and the colour match in meaning (stimulus-compatible) and are also associated with the same response key (response-compatible; e.g., the word “blue” printed in blue). Second, there are *same-response trials*, in which the word and colour are incompatible in meaning (stimulus-incompatible) but correspond to the same response key (response-compatible; e.g., the word “green” printed in blue, while “blue” and “green” responses are assigned to the same key). A difference between identity and same response trials indicates stimulus conflict. Third, there are *different-response trials* in which the word and colour are incompatible in meaning (stimulus-incompatible) and are also assigned to different response keys (response-incompatible; e.g., the word “red” printed in blue, while “blue” and “red” responses are assigned to different keys). A difference between same and different response trials therefore indicates response conflict (De Houwer, 2003; Hasshim & Parris, 2014; Jongen & Jonkman, 2008).

Comparisons with these three trial types allows for measures of both stimulus conflict (controlling for response conflict) and response conflict (controlling for stimulus conflict). In particular, identity and same response trials are both response compatible (i.e., both the colour and word suggest the same response key), but they differ in stimulus compatibility. Past research has shown that participants respond faster to identity trials relative to the same response trials (De Houwer, 2003; Schmidt & Cheesman, 2005). In the literature, the *stimulus conflict* effect is generally thought to result from semantic conflict (i.e., conflict in meaning between the word and colour). In addition, both same and different response trials are stimulus incompatible (i.e., the word and colour mismatch in meaning), and differ in response compatibility (i.e., the word and colour suggest the same key response on the same response trials, but different responses on different response trials). Past research has shown that same response trials are responded to faster and more accurately than different response trials (De

Houwer, 2003; Schmidt & Cheesman, 2005). This response conflict effect is normally interpreted as evidence for response competition.

Prior studies have shown that not all types of stimuli produce both stimulus and response conflict. Colour associates (i.e., words that are related in meaning to colours; e.g., “fire,” which is associated with red) also produce a congruency effect, which has been explained in terms of different mechanisms across several studies (Klein, 1964; Sharma & McKenna, 1998). Like colour words, colour associates can be either congruent (e.g., “fire” in red) or incongruent (e.g., “fire” in blue) with the associated ink colour. Early research described the Stroop effect for colour associates alternatively as due to stimulus or response conflict. A study conducted by Schmidt and Cheesman (2005) using the 2-to-1 mapping procedure found exclusively stimulus conflict for colour associates. This was interpreted as indicating that the colour associate can facilitate and/or interfere with processing of the ink colour on the semantic level. However, the relationship between the colour associate and compatible colour concept might not be strong enough to bias a potential response (e.g., “fire” facilitating “red” strong enough to retrieve a response linked to “red”). Thus, colour associates yield only the difference between identity and same response trials (stimulus conflict), but not the difference between the same-response and different-response trials (response conflict). In simple contingency learning experiments, in contrast, only a response conflict effect is observed (Schmidt et al., 2007; Schmidt & De Houwer, 2012). With matching-to-sample training, which is at least analogically similar to language learning (e.g., flash cards), a Stroop interference effect is induced and it is driven by response conflict only (Liefooghe et al., 2020).

The Stroop task has also been used to study language interference in bilinguals. The effect has been evidenced with both colour-word distracters of the first language (L1) and the second language (L2). For example, a native French speaker who also speaks English as a second language will be impaired by both English and French incongruent colour words. The standard finding is that the congruency effect is smaller for L2 than for L1 words. For instance, “rouge” (French for “red”) printed in blue will be more impairing for a native French speaker than “red” printed in blue. This congruency effect has been evidenced even for recently learned second language words. Altarriba and Mathis (1997) trained English monolinguals with a set of Spanish colour words. After training, participants named the colour of English and Spanish colour word distracters in English. As expected, the congruency effect was observed for both English and Spanish distracters. However, the congruency effect was smaller for Spanish colour words. These results suggest that even recently learned incongruent distracter words can slow down colour naming.

In another study with native German speakers, to-be-learned pseudowords were paired with German (L1) colour words (Geukes et al., 2015). In the training phase, participants learned an association between a given pseudoword and its assigned German “translation”. Pseudowords and German colour words were introduced in the Stroop task performed immediately after the learning phase or after 24 hours. Interestingly, the Stroop effect occurred for novel-words immediately after learning if they were presented together with the German colour-words. These findings suggest the automatic availability of the novel, recently trained words. The authors further suggested that the pseudowords have access to semantics, though a dissociation procedure was not used to separate stimulus and response conflict.

Prior studies such as these have established the presence and size of the congruency effect in native and second languages but have not investigated the source of this conflict. The first study that investigated whether the second language interference effect is due to stimulus conflict, response conflict, or both, was done by Schmidt, Hartsuiker, and De Houwer (2018). Using a Dutch-French version of the Stroop task with the same 2-to-1 mapping procedure, they observed both stimulus and response conflict in the first (i.e., Dutch) and second (i.e., French) language. Thus, even though the Dutch bilingual participants had relatively weak French skills, both stimulus and response L2 interference effects were evidenced. On the other hand, the French level of participants in this sample, though low, was more advanced than a beginner level, which is relevant for the present research question, discussed shortly.

The focus of the current study is to determine the source of the conflict within the system for a weakly spoken foreign language: at the level of semantics (stimulus conflict), at the level of responses (response conflict), or both. This will provide new perspective on the connection between L1 and L2 lexicons and semantic representations early in L2 learning. For instance, the *Revised Hierarchical Model* (Kroll & Stewart, 1994), presented in Figure 2, assumes that: a) lexical representations of L1 and L2 words are connected with a strong lexical link between L2 words and their L1 equivalents (e.g., for native Croatian speaker, “vert” is learned as a direct translation of “zelena”), and b) L1 words are more strongly connected to semantics (e.g., green colour) relative to L2 words.

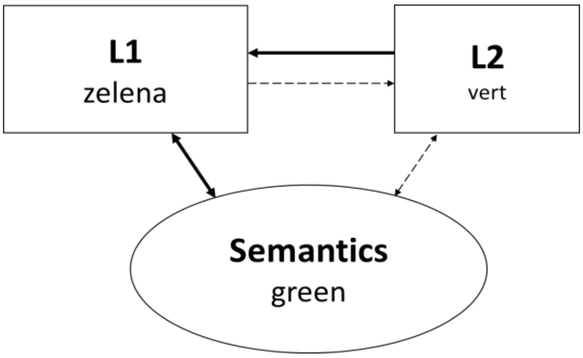

Fig. 2. Revised Hierarchical Model of Kroll and Stewart (1994).

Note. The model represents the connections between two lexicons and semantics. The asymmetries in the strengths of certain connections can be observed. Dashed lines = weak connections, solid lines = strong connections.

Kroll and Stewart’s model (1994) assumes that direct links between semantics and the L2 lexicon are initially very weak. As such, L2 colour words should not produce stimulus conflict early on in training. Even though the presence of stimulus interference effects was not examined directly, Altarriba and Mathis (1997) argued that conceptual links are formed for second language words, even after very short learning session (in contrast to Revised Hierarchical Model). According to this account, the word “rouge” printed in yellow activates the semantic representation for red as well as the semantic representation of yellow, resulting in semantic conflict. However, “red” does not interfere with response selection. The Revised Hierarchical model also assumes that L2 connections to semantics should emerge eventually. Thus, as a potential caveat for the results of Schmidt and colleagues (2018), participants may have been fluent *enough*: French fluency in the Flemish region of Belgium, though not substantial, was perhaps strong enough to induce both stimulus and response conflict. Taken together, there are certain indications suggesting that L2 words influence cognitive processing similarly to L1 words, but other theories suggest fundamental differences (e.g., Kroll and Stewart’s model).

In the present series of experiments, we used the 2-to-1 mapping Stroop design to dissociate stimulus and response conflict. Experiment 1 was similar in concept to the Dutch-French study of Schmidt and colleagues (2018), except that we aimed to use a second language that was even less fluent (English in native French speakers). We anticipated that English fluency in France would prove much weaker than French fluency in Flanders. In Experiment 2, to see whether the link between L2 lexicon and semantics could occur even at very early stages of novel language acquisition, we trained participants with a completely novel, obscure language (Croatian, with native English speakers). Thus, Experiment 2 was similar in concept to the study of Altarriba and Matthis (1997), but with the addition of a 2-to-1 response mapping.

Three possible patterns of results might occur. Firstly, both stimulus and response conflict might be observed for L2 words, as is the case for L1 words (i.e., identity < same response < different response). Since conflict effects are generally smaller for L2 than L1 words, the assumption that one or even both conflict components are decreased for L2 words seems reasonable. The second possibility is that foreign colour words produce exclusively stimulus conflict. In this case, L2 words act in similar fashion as L1 colour associates (i.e., “fire” as an associate to “red”). Therefore, only a difference between identity and same response trials is expected, not the difference between same and different response trials. The third possibility is that L2 words might produce exclusively response conflict. In other words, an L2 distracter word biases response selection without interfering with stimulus colour identification, as suggested by the Kroll and Stewart’s (1994) model.

In order to exclude the possibility that stimulus or response conflict (or both) for L2 words might be produced by the similarity between colour words across two languages (e.g., “bleu” in French and “blue” in English), we selected colour word pairs (i.e., French-English in Experiment 1 and English-Croatian in Experiment 2) which do not share an etymology. Cognates are the words that are similar across languages due to a shared etymology (e.g., “blue” and “bleu”). Non-cognates are dissimilar (e.g., “white” in English and “blanc” in French), with little or no overlap in spelling or pronunciation. Larger interference effect are found for cognates (Dyer, 1971; van Heuven et al., 2011) due to spelling similarities (priming), so for the purposes of the present studies, we opted to use French-English and English-Croatian non-cognates.

The similarities in word processing between two languages and their link to semantics have been a center of much debate so far. The question of whether the L2 distracting words elicit both stimulus and response conflict has some implications for theorizing about language cognition. The connection between first and second languages seems to influence semantic identification and response decision processes while performing a bilingual colour-word Stroop task. There is already some evidence that the L1 and L2 (learned early in school) colour-words produce similar pattern of impairment in semantic and response processing in the Stroop task (Schmidt et al., 2018). In the present manuscript, we aimed to investigate whether the same pattern could be observed with even lower levels of L2 proficiency.

2. Experiment 1

In Experiment 1, we investigate the source of conflict in a very low fluency second language. By using a French-English version of the Stroop task, we aimed to examine whether low English proficiency can produce L2 stimulus and/or response interference.

2.1. Method

2.1.1. Participants

Eighty-five University of Burgundy undergraduates (72 female, 13 male) participated in the study. Participants were recruited by signing up on a sheet posted on the psychology department board. All participants had normal or corrected-to-normal colour vision and spoke French as a first language. They received course credit in exchange for participation.

2.1.2. Apparatus and materials

Stimuli were presented on a standard 15’’ PC laptop. Stimulus presentation and response collection were controlled by E-prime 2.0 software. Responses were made on a standard AZERTY keyboard, with the ‟F” (left) and ‟J” (right) keys. Prior to the computer portion of the experiment, participants were given a pen-and-paper survey to fill out. The first part of the survey was the LEXTALE_EN (Lemhöfer & Broersma, 2012), with instructions translated into French. Within this test, 84 English-looking words are presented. About 2/3 of the presented words are actual English words (e.g., “denial”), whereas the remaining 1/3 are not (e.g., “platery”). Participants were informed to select the words that they are certain are actual English words. Correct answers were rewarded with one point, and incorrect trials were penalized by two points. The questionnaire also asked for gender, native language, years of English study in school, and a self-rating of English knowledge on a scale from 0 (“almost none”) to 10 (“perfect”). After this, a subset of questions from the French version of the

Language Experience and Proficiency Questionnaire – LEAP-Q (Marian et al., 2007) were appended. The first three questions from the Questionnaire were retained, which asked, respectively, for a list of languages in order of dominance, a list of languages in order of acquisition, and the percentage with which the participant used each of their spoken languages in the recent period. Also retained from the LEAP-Q were two boxes, one for French and another for English, asking for the age that the participant began acquiring the language, became fluent in the language, began learning to read in the language, and became fluent in reading the language. The purpose of these questions was to assure participants had the correct language dominance. These metrics were also correlated with the observed congruency effects to determine whether any observed congruency effects were strongly dependent on fluency. Finally, participants were asked to give the English translations of the four French colour words used in the experiment. The purpose of this question was to see how familiar the stimuli were to participants and to assure they knew the correct translation of each of the colour words.

2.1.3. Design

In the main part of the experiment (i.e., after the survey), participants were presented with the French and English words for "green", "yellow", "silver", and "pink" (in French: "vert", "jaune", "argent", and "rose", respectively). We selected these four pairs of colour words because they are non-cognates (unlike some other colour words: e.g., "blue/bleu" or "red/rouge"). The print colours were green (0,128,0), yellow (255,255,0), silver (192,192,192), and pink (255,105,180), corresponding to "green", "yellow" "silver", and "hotpink" in the standard E-prime/HTML colour palette. For each participant, two colours were mapped to the left key ("F") and another two to the right key ("J"). The combinations of the colours mapped to each key were fully counterbalanced across participants (six factorial combinations).

Two factors were manipulated in the within-subject design. The first factor was the distracter language (French vs. English) and the second was congruency (identity – the word and the print colour match; same response – the word and the print colour mismatch but are mapped to the same key; different response – the word and the print colour mismatch and are mapped to different keys).

The study consists of one practice block and three main experimental blocks. The experimental blocks were separated by a five-second pause. The practice block had 64 trials. Within the practice block, the stimulus "xxxx" was presented in lowercase 16 times in each colour. In each of the experimental blocks, there were two sub-blocks in which each of the eight colour words was presented once in all four colours (i.e., 32 trials per sub-block, 64 in total) selected randomly without replacement. There were therefore 192 experimental trials across the three experimental blocks.

2.1.4. Procedure

Participants sat approximately 60 cm away from the screen. They were asked to read carefully the instructions presented on the screen, place their fingers on the "F" and "J" keys and to respond as fast as possible without making too many errors. Stimuli were presented on a black screen in 18 pt., bold Courier New font. Each trial started with the fixation ("+") presented in the center of the screen for 250 ms. This was followed by a blank screen for 250 ms. The coloured word/letter string was then presented in the center of the screen until a response was registered or 2000 ms elapsed. If the participant made an error or failed to respond within 2000 ms, then the message "Erreur" ("Error/Incorrect") or "Trop Lent" ("Too slow"), respectively, appeared in red for 1000 ms before the next trial.

2.2. Results

Firstly, we analyzed average language metric scores, which are presented in Table 1. Most participants rated English as their second most dominant and second acquired language. Brief inspection of the age of gaining English language skills (speaking, reading) suggest they started quite late (9-15.5 years). Despite learning English for more than 9 years on average, participants self-rated their English proficiency relatively moderately (5.50 on 1-10 scale), and the objective English vocabulary knowledge scores (67.01) were quite low. Participants mainly use French in their everyday life (78.25%), and they have relatively little exposure to English (13.69%). More information on the language demographics of the sample are available in the Appendix.

Table 1

Mean language scores and standard errors.

	Mean	SE
LEXTALE_EN		
Years English	9.65	.247
English Level	5.50	.191
Score	67.01	.948
LEAP-Q		
Dominance French	1	0
Dominance English	2.35	.065
Order French	1.01	.012
Order English	2.32	.056
French Use (%)	78.25	1.722
English Use (%)	13.69	1.139
French		
Acquisition	1.53 years	.154
Fluent	3.53 years	.222
Reading	5.39 years	.153
Fluent Read	7.44 years	.317
English		
Acquisition	9.08 years	.226
Fluent	14.48 years	.462
Reading	12.11 years	.382
Fluent Read	15.57 years	.551

Secondly, we analyzed mean correct response time and percentage error data of the Stroop task. For response times, only correct responses were assessed. For error percentages, any responses above 2000 ms were considered spoiled trials and were excluded from analysis. As a supplementary analysis, we included Bayesian statistics using the standard noninformative Cauchy prior in JASP (Marsman & Wagenmakers, 2017) with a default width of 0.707.

2.2.1. *Response time*

To analyze response times, we conducted a congruency (identity vs. same response vs. different response) by language (French vs. English) within-subjects repeated measures ANOVA. The correct response time data are shown in Figure 3. There was a main effect of congruency, $F(2,168) = 36.48$, $p < .001$, $MSE = 1581.64$, $\eta_p^2 = .30$, $BF_{10} = 9.778e^{11}$. However, the main effect of language was not significant, $F(1,84) = .81$, $p > .05$, $MSE = 1657.54$, $\eta_p^2 = .01$, $BF_{10} = .139$. The interaction between congruency and language was only marginally significant, $F(2,168) = 2.67$, $p = .07$, $MSE = 1637.75$, $\eta_p^2 = .03$, $BF_{10} = 46.25$.

The comparison of response times between different types of trials was conducted separately for French and English words. For French colour words, we found significant stimulus conflict (same response – identity), $t(84) = 4.77$, $p < .001$, $MEAN_{diff} = 34.06$, $SE_{diff} = 7.14$, *Cohen's d* = .52, $BF_{10} = 2142.21$. Surprisingly, the response conflict (different response – same response) effect failed to reach significance, $t(84) = 1.25$, $p = .21$, $MEAN_{diff} = 8.17$, $SE_{diff} = 6.52$, *Cohen's d* = .14, $BF_{10} = .254$. For English colour words, both stimulus conflict, $t(84) = 2.28$, $p < .05$, $MEAN_{diff} = 13.89$, $SE_{diff} = 6.10$, *Cohen's d* = .25, $BF_{10} = 1.375$ and response conflict, $t(84) = 3.01$, $p < .01$, $MEAN_{diff} = 16.31$, $SE_{diff} = 5.41$, *Cohen's d* = .33, $BF_{10} = 7.853$, were significant. The magnitude of the stimulus conflict effect was larger for French than for English words, $t(84) = 2.20$, $p < .05$, $MEAN_{diff} = 20.17$, $SE_{diff} = 9.17$, *Cohen's d* = .24, $BF_{10} = 1.173$, but there was no difference in the magnitude of the response conflict effect across languages, $t(84) = .88$, $p > .05$, $MEAN_{diff} = -8.14$, $SE_{diff} = 9.27$, *Cohen's d* = -.09, $BF_{10} = .174$.

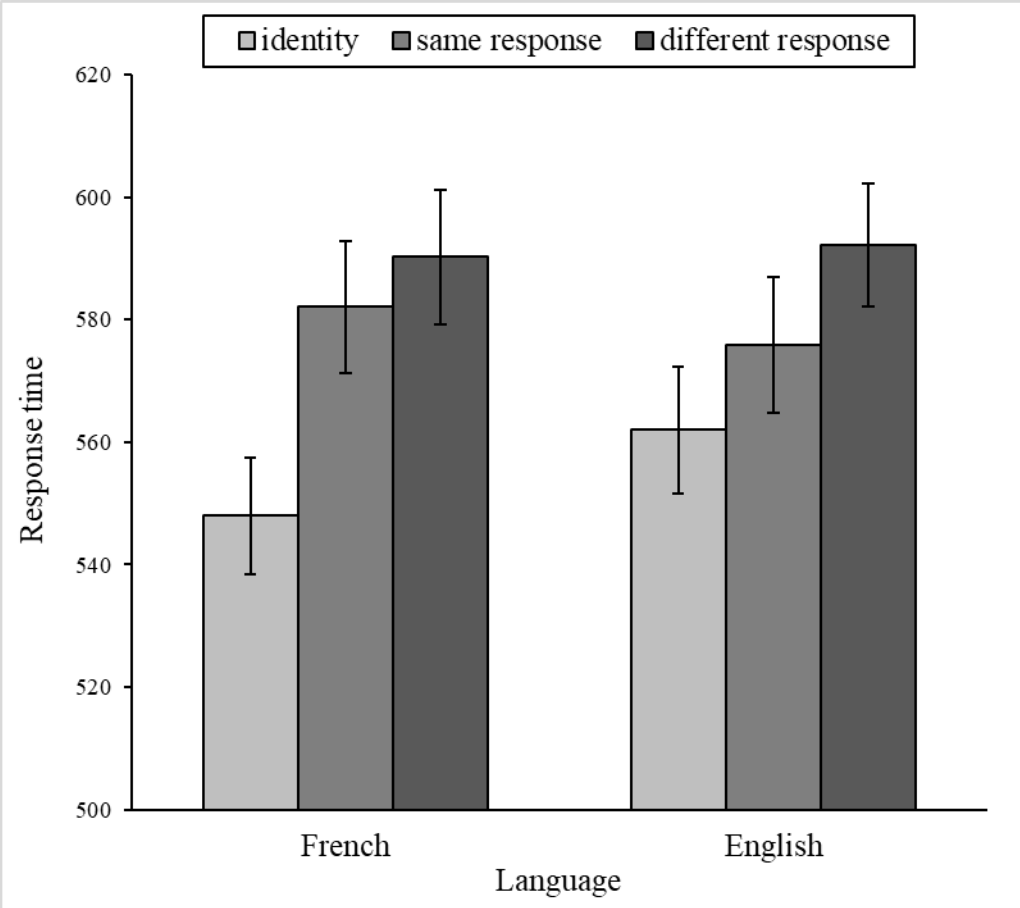

Fig. 3. Response times (in milliseconds) for French and English colour words in the Stroop task. Error bars depict standard error of means.

2.2.2. Percentage error

As with the response time data, we again conducted a congruency (identity vs. same response vs. different response) by language (French vs. English) within-subjects repeated measures ANOVA. The mean percentage error data are presented in Figure 4. The main effect of congruency was significant, $F(2,168) = 12.45$, $p < .001$, $MSE = 19.70$, $\eta_p^2 = .13$, $BF_{10} =$ 11076.42, but there was no main effect of language, $F(1,84) = .14$, $p > .05$, $MSE = 16.57$, $\eta_p^2 = .002$, $BF_{10} = .104$. The interaction between congruency and language was not significant, $F(2,168) = 1.71$, $p > .05$, $MSE = 15.08$, $\eta_p^2 = .02$, $BF_{10} = .18$.

Further comparisons between different types of trial were conducted on each language separately. For French colour words, there was no stimulus conflict effect, $t(84) = .32$, $p >$

.05, $MEAN_{diff} = -.19$, $SE_{diff} = .58$, *Cohen's d* = .03, $BF_{10} = .123$, but there was a significant response conflict effect, $t(84) = 4.51$, p < .001, $MEAN_{diff} = 2.74$, $SE_{diff} = .61$, *Cohen's d* = .49, $BF_{10} = 3.202e^{13}$. Similarly for English colour words, the stimulus conflict effect was not significant, $t(84) = 1.17$, $p > .05$, $MEAN_{diff} = -.83$, $SEdiff = .71$, *Cohen's d* = .13, $BF_{10} = .276$, but the response conflict effect was significant, $t(84) = 2.77$, $p = .01$, $MEAN_{diff} = 1.83$, $SE_{diff} = .66$, *Cohen's d* = .30, $BF_{10} = 3.296e^{10}$. There was no evidence for any differences in the magnitude of the stimulus conflict effect, $t(84) = .71$, $p > .05$, $MEAN_{diff} = .65$, $SE_{diff} = .90$, *Cohen's d* = .08, $BF_{10} = .153$, nor response conflict effect $t(84) = 1.16$, $p > .05$, $MEAN_{diff} = .90$, $SE_{diff} = .78$, *Cohen's d* = .13, $BF_{10} = .229$ across languages.

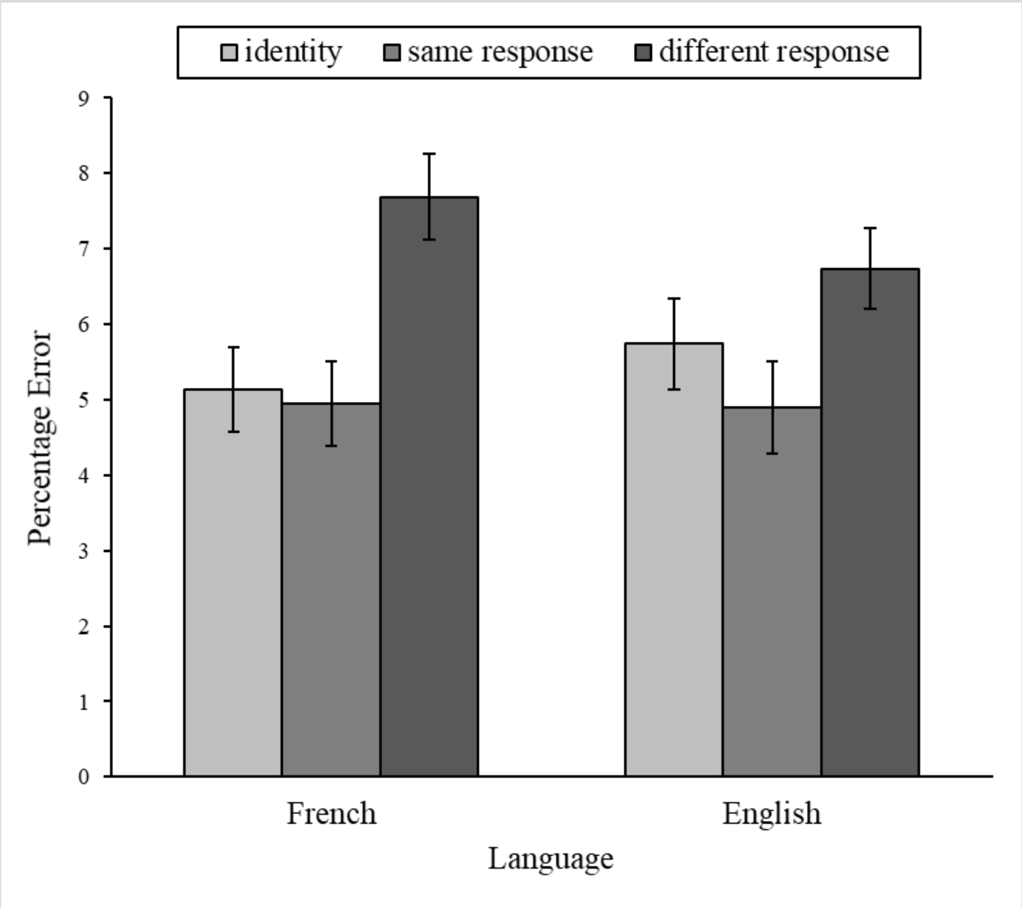

Fig. 4. Percentage errors for French and English colour words in the Stroop task. Error bars depict standard error of means.

2.2.3. Correlations

Additionally, we assessed the level to which language-related variables correlate with the stimulus and response conflict effects for both French and English colour words. The non-parametric rank-based Spearman's *ρ* correlation coefficients are displayed in the Table 2. As seen from the table, none of the behavioral measures (French or English stimulus or response conflict effect) correlate with the self-rated English proficiency level. Considering the LEAP-Q variables, the percentage of English language use, age of French reading acquisition, and age of French and English fluent reading did not correlate with behavioral Stroop measures. After applying a Holm-Bonferroni correction for multiple comparisons, none of the correlations were significant at *a* = .05.

Table 2

Correlations of demographic variables with conflict effects.

	French				English			
	Stimulus		Response		Stimulus		Response	
	RT	ERR	RT	ERR	RT	ERR	RT	ERR
LEXTALE_EN								
Years English	.003	*-.233*	-.013	*.266*	-.141	-.139	**.311**	.173
English Level	-.019	-.068	-.018	.002	-.085	.047	.049	.022
Score	-.089	.038	.081	-.027	-.110	-.131	.105	*.230*
LEAP-Q								
% French Use	.001	.048	-.006	-.024	*-.242*	.035	.134	-.115
% English Use	-.105	-.095	-.044	.046	.096	.098	.008	-.011
French								
Acquisition	.033	-.164	-.055	.138	-.176	*.237*	-.038	-.165
Fluent	-.009	-.112	-.176	.158	-.120	**.312**	.048	-.176
Reading	.099	-.051	-.117	-.157	-.130	.067	.159	-.106
Fluent Read	.008	.122	-.145	-.156	-.067	-.091	.054	.039
English								
Acquisition	-.137	.172	.101	-.206	.089	.000	*-.239*	-.069
Fluent	*-.299*	-.118	.149	-.004	.004	.110	.065	-.062
Reading	.128	-.002	-.132	-.007	**.327**	-.040	*-.252*	-.123
Fluent Read	-.169	.056	.105	.076	.117	-.122	-.028	-.002

Note. Bold = p<.01, Italic = p<.05. No correlation is significant after Holm-Bonferroni correction.

2.3. Discussion

In Experiment 1, we investigated the source of congruency effects in a weakly spoken second language with 2-to-1 keypress mapping assignment. In line with the previous study with a more fluent second language (Schmidt et al., 2018), both stimulus and response conflict effects were found for a second language (English). Unexpectedly, only the stimulus conflict effect, but not the response conflict effect, was significant for the first language (French). This is almost certainly a Type 2 error given that response conflict has been observed repeatedly in L1. Although non-significant, response conflict in L1 was numerically in the expected direction with slower response latencies for different response trials relative to same response trials. Further, the response conflict effect was large and robust in the errors. This might suggest a speed-accuracy trade-off (or simply a Type 2 error in response times). We also note that the absence of a stimulus conflict effect in the errors for both languages is not unexpected, as stimulus conflict is almost never present in the error data with this design.

More important, of course, are the L2 results. Our L2 results contrast the hypothesis that second language words produce exclusively stimulus conflict (Glaser & Glaser, 1989; MacKinnon et al., 1985) or exclusively response conflict (Klein, 1964; Sharma & McKenna, 1998). Our participants showed a relatively poor English level, as assessed by self-report and LEXTALE vocabulary test scores, but it seems that English colour words affect stimulus and response processing in similar fashion as the French colour words. The observed conflict effects also did not seem to correlate strongly with measures of fluency. Despite this low English level in participants within the sample, one potential limitation with Experiment 1 is that French participants are certainly familiar with English colour words. It could be that during very early language acquisition only stimulus or only response conflict are present, but that our participants were sufficiently familiar with the colour words used in our study to produce both. In Experiment 2, we address this potential caveat by training participants with completely unfamiliar colour words.

3. Experiment 2

From the perspective of Kroll and Stewart's model (1994), novel foreign words should be very weakly connected to semantics. One of the basic assumptions of the model is that at early stages of foreign language acquisition, only a lexical link between a novel word and its first language translation will be established. In other words, we memorize a novel word as a direct translation of its first language equivalent. However, the connections between the L2 lexicon and semantics will emerge eventually with increased second language proficiency (Sholl et al., 1995; Talamas et al., 1999). Results of Experiment 1, similar to those of Schmidt and colleagues (2018), indicate that L2 words (even when fluency is very low) influence semantic and response processing in a similar way as L1 words. As a further test of this notion, we aimed to investigate whether the same interference pattern can be observed for the words from a completely unknown language after brief language training.

To test this assumption, we take the logic of Experiment 1 even further by training participants with completely unfamiliar colour words. In this way, there can be no doubt that we are studying very early language acquisition. In particular, we introduced a short and simple learning and training procedure (approximately 15-20 minutes), in which participants were first exposed to English-Croatian colour word pairs (passive learning procedure). In the following training phase, they were encouraged to strengthen the link between word equivalents, similarly as in previous studies (Altarriba & Basnight-Brown, 2012; Altarriba & Mathis, 1997) but with a test phase to separate stimulus and response conflict.

In interest of full disclosure, we note that we initially ran two other studies with University of Burgundy undergraduate students, which included a notably shorter training procedure[1]. Unfortunately, these studies did not induce a meaningfully large Croatian Stroop

[1] In the first pilot study, the training phase was much shorter: it consisted of 32 trials. On each trial, a target word was presented together with four colour word labels (i.e., potential responses) printed in the corresponding ink colour. There were no catch trials in the Stroop task. The second pilot study consisted of 64

effect (essentially, a failed manipulation check), which made decomposition of the effect into stimulus and response conflict severely underpowered. We note that, visually, results were similar to the results of our final study reported here, but effects were much smaller. More details on the prior two studies can be obtained upon request.

3.1. Method

3.1.1. Participants

A total of 122 participants (85 women and 37 men) were recruited online via the prolific.ac website. An additional six submissions were rejected by the experimenter due to technical issues or inappropriate completion time (too short or too long). A further 62 submissions were incomplete, and therefore excluded. Over half of these dropouts (34) did not complete the survey portion of the task. We do not know how far the remaining dropouts (28) progressed in the task, as the Psytoolkit server does not store incomplete data (e.g., it is possible that a technical error occurred when launching the experimental portion of the task). We note the dropout rates like this are typical for online data collection, but results are generally very similar to lab-collected data (Crump et al., 2013). The selection of the final sample was based on the following criteria: speaking English as a native language, having normal or corrected-to-normal colour vision, and having no prior knowledge of the Croatian language. More demographic information about the sample can be found in the Results section. Participants were paid £3.5 for participation in the experiment, which lasted approximately 30 minutes.

3.1.2. Apparatus and materials

The experiment was programmed in the Psytoolkit (Stoet, 2010, 2017) programming language and designed to work on a PC. Prior to the experimental portion of the experiment, participants completed a short survey concerning their language background. The

training trials; the types of trials were identical to those in the present Experiment 2, but with four response alternatives. The Stroop task was identical to that in Experiment 2 of the present manuscript.

questionnaire asked for gender, age, and native language. To make sure participants had no prior knowledge of the Croatian language, we asked whether they have ever studied Croatian in school, to self-rate their Croatian knowledge from 1 (“not at all”) to 10 (“very well”), and to translate the four Croatian colour words used in the experiment (i.e., “crvena”, “plava”, “zelena”, and “siva”). After this, there was a subset of questions taken from the English version of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007), the identical questions as we used in Experiment 1. The purpose of the survey portion of the study was to assure that all participants had the target language dominance (i.e., they are native English speakers) and no previous experience with the Croatian language.

3.1.3. Design

3.1.3.1. Learning phase. In the learning phase, participants were presented with the Croatian and English colour words for “red”, “blue”, “green”, and “gray” (in Croatian “crvena”, “plava”, “zelena”, and “siva”, respectively). The learning phase consisted of one block of four Croatian-English word pairs presented four times each, in randomized order. Participants were asked to pay attention on word pairs and try to memorize them.

3.1.3.2. Training phase. The training phase started immediately after the learning phase. There were four types of trials in the Training phase: 1) Croatian target with English labels, 2) Croatian target with colour box labels, 3) English target with Croatian labels, and 4) Colour box target with Croatian labels. These trials are illustrated in Figure 5. There were 144 trials for each trial type, for a total of 576 training trials. Participants responded by pressing the F-key for the left response option and the J-key for the right response option. They were instructed to respond as quickly and accurately as possible.

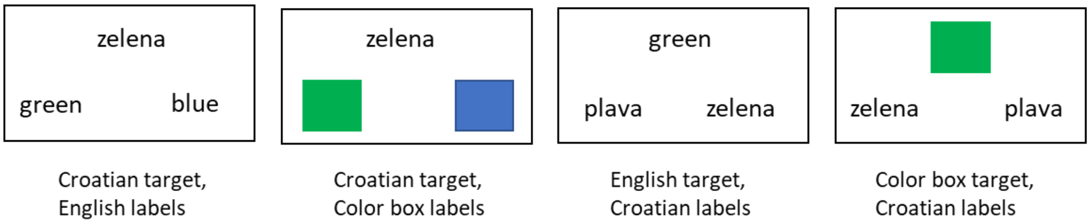

Fig. 5. Illustration of four types of trial presented in the Training phase. See the online version of this article for colour figures.

3.1.3.3. Stroop task. The final Stroop phase of Experiment 2 was identical in all respects to Experiment 1, with a few minor exceptions. First, the word and colour stimuli were changed. In particular, the Croatian words were “crvena”, “plava”, “zelena”, and “siva”. The corresponding English words and the RGB values for the stimulus colors were “red” (255, 0, 0), “blue” (0, 0, 255), “green” (0, 100, 0), and “gray” (128, 128, 128). The practice block was also identical, except that there were 72 practice trials. Each experimental block consisted of 72 trials (in total 216). Unlike Experiment 1, approximately 20% of experimental trials were “catch” trials. On these catch trials, a novel Croatian word was presented (“mokar” or “nakon”) in one of four colours. Participants were instructed to withhold their response on these trials. Participants were told in advance the two catch words but were not informed of their meaning. This manipulation was used to increase the Stroop effect. It requires participants to attend to the distracter words (i.e., to detect whether either of the two catch words was presented), thereby increasing the influence of the words on responding (Eidels et al., 2014).

3.1.4. Procedure

After completing the language background questionnaire, the main experiment began. It consisted of 3 phases: Learning phase, Training phase, and Stroop task.

3.1.4.1. Learning phase. Stimuli were presented on a white screen in 56 pt. Courier New font. On each trial, participants were first presented with a fixation “+” in black for 1000 ms. Next, a coloured Croatian word appeared in the center of the screen. After 1000 ms, the English translation printed in the corresponding colour appeared below it. Both words remained on the screen for 4000 ms.

3.1.4.2. Training phase. On each trial, there was a fixation point displayed in the center of the screen for 500 ms. Next, the target appeared in the center of the screen together with two labels representing two response options located below the target on the left (-200, 200) and on the right (200, 200). The locations of the response options were randomized from one trial to the next. The target and the labels remained on the screen for 3000 ms or until a response was registered. The next trial began immediately following a correct response. If the participant made an error or failed to respond in 3000 ms, the message “Incorrect” or “Too slow”, respectively, appeared in black for 1000 ms before the next trial.

3.1.4.3. Stroop task. On each trial of the Stroop task, participants were first presented with a fixation “+” for 250 ms. This was followed by a blank screen for 250 ms. Afterwards, the coloured word was displayed on the screen until a response was registered or 2000 ms elapsed. The same error and time-out messages as the prior phase were used for incorrect responses and trials where participants failed to respond before the 2000 ms deadline. Catch trials were presented for a fixed duration of 750 ms. If participants responded before this, the message “Do not respond to this word” appeared and remained on the screen for 1000 ms.

3.2. Results

Firstly, we analyzed the language demographic data based on the responses in the survey portion of the study. Almost all participants indicated English as their first dominant language (95%), and as their first language in order of acquisition (98%). The most frequent second language in order of dominance was French (30%), followed by German (18%) and Spanish (16%). In order of acquisition, French was rated as a second language by 41% of participants, followed by Spanish (15%). Other languages, such as Japanese, Hindi, Punjabi, Turkish, Italian, Irish, and Welsh were also noted as second languages both in order of dominance and acquisition in small percentages. None of participants studied Croatian in school. All of them self-rated their Croatian knowledge as 1 (“not at all”) and were unable to translate the Croatian colour words. Secondly, we assessed mean correct response time and percentage error for the Training phase and the Stroop task.

3.2.1. Training phase

All the participants had reasonable percentage errors in the Training phase (<20%), so none were excluded from further analysis. Mean percentage error was 4.25% (SE = .29). Means and standard errors for response times and percentage errors across the four types of trials presented in the Training phase are displayed in the Table 3.

Table 3

Training phase means and standard errors.

Type of trial	Response time		Percentage error	
	Mean	SE	Mean	SE
Croatian target, English labels	1118.19	13.25	6.03	0.38
Croatian target, colour box labels	969.14	12.47	3.99	0.3
English target, Croatian labels	1079.87	12.99	3.3	0.29
Colour box target, Croatian labels	1062.24	13.39	3.67	0.31

Participants responded significantly faster on trials with a Croatian target and colour box labels than to other types of trials: a) Croatian target with English labels, $t(121) = 24.37$, $p < .001$, $MEAN_{diff} = 149.05$, $SE_{diff} = 6.12$, $Cohen's\ d = 2.21$, $BF_{10} = 1.102e^{45}$, b) English target with Croatian labels, $t(121) = 17.23$, $p < .001$, $MEAN_{diff} = 110.73$, $SE_{diff} = 6.43$, $Cohen's\ d = 1.56$ $BF_{10} = 1.213e^{31}$ and c) Colour box target with Croatian labels, $t(121) = 17.27$, $p < .001$, $MEAN_{diff} = 93.10$, $SE_{diff} = 5.39$, $Cohen's\ d = 1.56$, $BF_{10} = 1.445e^{31}$. No differences were observed between the remaining three conditions.

3.2.2. Stroop task

The performance on catch trials was analyzed separately from experimental trials within the Stroop tasks. On average, catch trials accounted for approximately 20% of experimental trials (*M*=42.97, *SE*=.13). Participants were relatively successful in withholding their responses on catch trials (*M*=83.83%, *SE*=1.83).

3.2.2.1. Response time. For response times, only correct responses were analyzed, with no other trims. Data are presented in the Figure 6. We conducted a congruency (identity vs. same response vs. different response) by language (English vs. Croatian) within-subject repeated measures ANOVA. The main effect of Congruency was significant; $F(2, 242) = 114.83$, $p < .001$, $MSE = 2826.67$, $\eta_p^2 = .49$, $BF_{10} = 3.147e^{36}$. The main effect of Language was marginally significant; $F(1,121) = 3.77$, $p = .054$, $MSE = 2906.55$, $\eta_p^2 = .03$, $BF_{10} = .297$. A significant Congruency by Language interaction was observed; $F(2, 242) = 13.16$, $p < .001$, $MSE = 3103.40$, $\eta_p^2 = .10$, $BF_{10} = 9269.47$.

Considering the significant interaction, we compared the different types of trials separately for each language. For English colour words, we observed a significant stimulus conflict effect (same response – identity); $t(121) = 7.33$, $p < .001$, $MEAN_{diff} = 53.5$, $SE_{diff} = 7.30$, *Cohen's d* = .66, $BF_{10} = 2.915e^{8}$, and response conflict effect (different response – same response); $t(121) = 6.39$, $p < .001$, $MEAN_{diff} = 44.4$, $SE_{diff} = 6.94$, *Cohen's d* = .579, $BF_{10} = 3.145e^{6}$. For Croatian colour words, stimulus conflict was significant; $t(121) = 4.96$, $p < .001$, $MEAN_{diff} = 35.2$, $SE_{diff} = 7.09$, *Cohen's d* = .449, $BF_{10} = 5833.98$; while response conflict showed a tendency toward statistical significance; $t(121) = 1.76$, $p = .08$, $MEAN_{diff} = 11.6$, $SE_{diff} = 6.60$, *Cohen's d* = .159, $BF_{10} = .449$. There was no difference in the magnitude of stimulus conflict between English and Croatian words; $t(121) = 1.80$, $p > .05$, $MEAN_{diff} = 18.3$, $SE_{diff} = 10.2$, *Cohen's d* = .163, $BF_{10} = .48$. However, the response conflict effect was larger for English words than for Croatian words; $t(121) = 3.19$, $p < .01$, $MEAN_{diff} = 32.8$, $SE_{diff} = 10.3$, *Cohen's d* = .289, $BF_{10} = 11.87$.

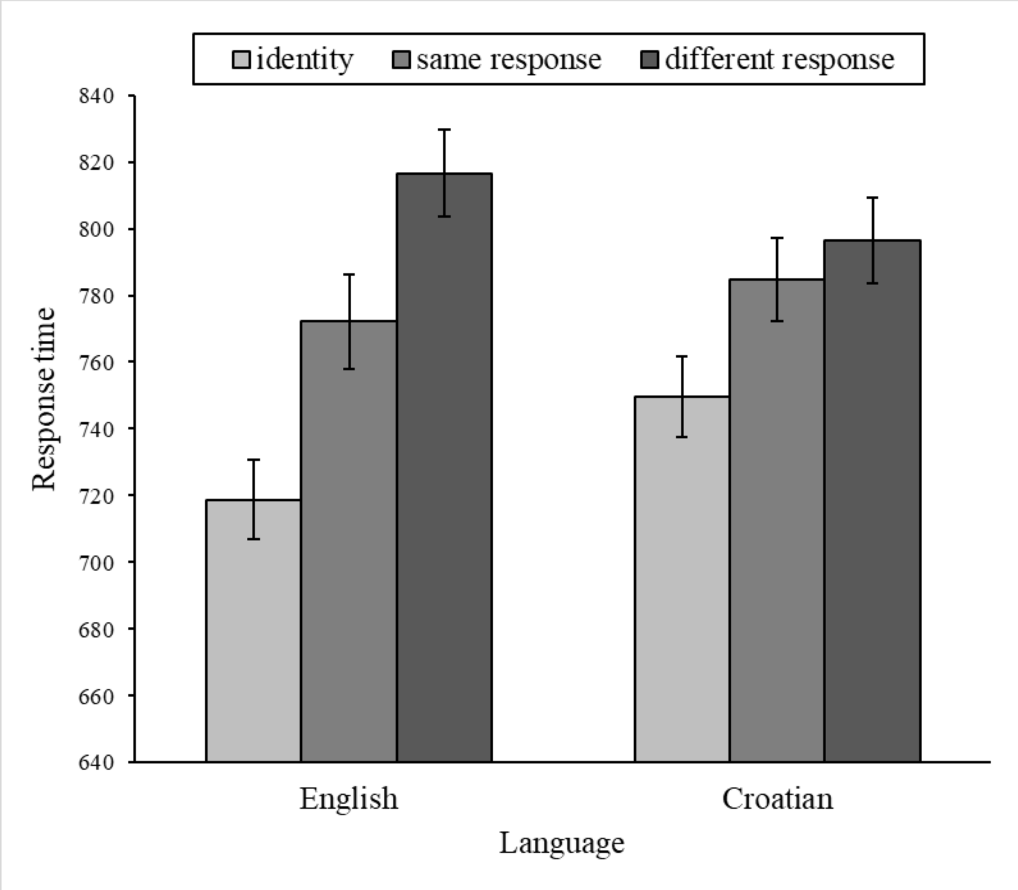

Fig. 6. Response times (in milliseconds) for English and Croatian colour words in the Stroop task. Error bars depict standard error of means.

3.2.2.2. Percentage error. For error percentages, trials in which participants failed to respond before the deadline were excluded (1.6% of trials). Percentage error data are presented in Figure 7. Again, a congruency (identity vs. same response vs. different response) by language (English vs. Croatian) within-subjects repeated measures ANOVA was conducted. The main effect of Congruency was observed; $F(2, 242) = 31.84$, $p < .001$, $MSE = 38.34$, $\eta_p^2 = .21$, $BF_{10} = 3.516e^{12}$. However, there was no main effect of language; $F(1,121) = 2.72$, $p > .05$, $MSE = 26.39$, $\eta_p^2 = .02$, $BF_{10} = .203$. The interaction between congruency and language was significant; $F(2, 242) = 4.63$, $p = .01$, $MSE = 32.00$, $\eta_p^2 = .04$, $BF_{10} = 1.80$.

We again compared performance on the different types of trials across languages. For English colour words, stimulus conflict was not significant; $t(121) = .31$, $p > .05$, $MEAN_{diff}$ =

.224, SE_{diff} = .727, *Cohen's d* = .028, BF_{10} = .105; but response conflict was significant; $t(121) = 7.30$, $p < .001$, $MEAN_{diff}$ = 5.11, SE_{diff} = .70, *Cohen's d* = .661, BF_{10} = 2.529e^{8}. Similarly for Croatian colour words, stimulus conflict was not significant; $t(121) = .09$, $p > .05$, $MEAN_{diff}$ = .069, SE_{diff} = .804, *Cohen's d* = .008, BF_{10} = .101; while response conflict was significant; $t(121) = 3.41$, $p = .001$, $MEAN_{diff}$ = 2.49, SE_{diff} = .729, *Cohen's d* = .309, BF_{10} = 23.22. No significant difference between English and Croatian stimulus conflict was observed; $t(121) = .16$, $p > .05$, $MEAN_{diff}$ = .155, SE_{diff} = .974, *Cohen's d* = .014, BF_{10} = .102. However, the response conflict effect was significantly larger for English words than for Croatian words; $t(121) = 2.49$, $p = .01$, $MEAN_{diff}$ = 2.62, SE_{diff} = 1.05, *Cohen's d* = .226, BF_{10} = 1.95, similar to the response times.

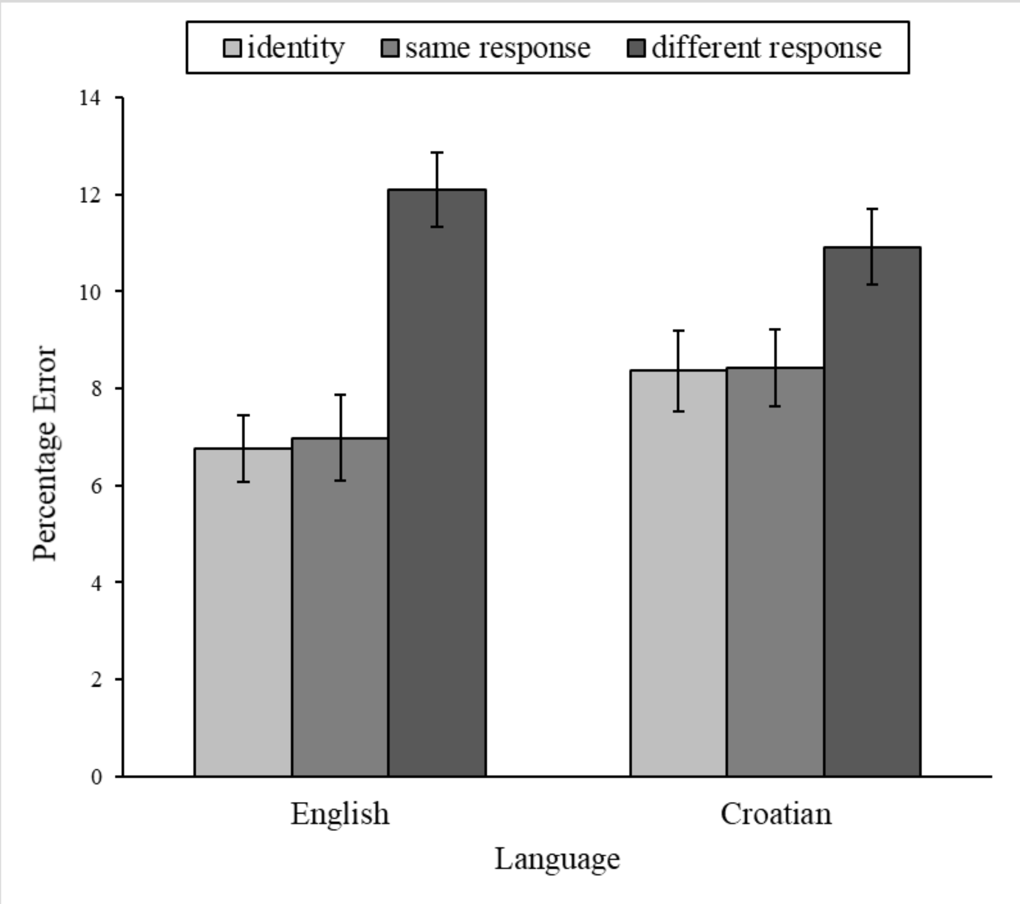

Fig. 7. Percentage errors for English and Croatian colour words in the Stroop task. Error bars depict standard error of means.

3.3. Discussion

Experiment 2 investigated the congruency effect in a recently trained, unfamiliar language. We used a relatively short (approximately 15-20 minutes) and simple training procedure in order to establish and reinforce a link between the Croatian words and their English equivalents. Interestingly, recently-trained Croatian words produced significant stimulus conflict in response times, similar to the L1 colour words. This stimulus conflict observed for Croatian words could be interpreted as a result of early, semantic processes (Glaser & Glaser, 1989; MacKinnon et al., 1985) in word acquisition. It seems plausible that two dimensions of a novel word (i.e., ink colour and word meaning) have been associatively related over the training phase. Consequently, on incongruent trials in the Stroop task, the concurrent activation of the colour and ink produce stimulus conflict. Our results suggest that, contrary to Revised Hierarchical model, even recently acquired novel words might link to semantics. As previously noted, stimulus conflict is rarely observed in the errors in the 2-to-1 procedure and was unsurprisingly not observed for either language in the current experiment (as in Experiment 1).

We also observed a response conflict effect for Croatian words that was significant in errors and marginal in response times. The presence of response conflict for the newly learned foreign colour words also argues against a model assuming exclusively semantic learning in early language acquisition. Interestingly, however, the response conflict effect was significantly reduced for L2 words relative to L1 words in both response times and errors. Although our main goal was to determine the presence or absence of stimulus and response conflict for L2 words, our results further suggest that response conflict is reduced in magnitude during early phases of L2 training.

As one potential caveat, the Stroop portion of the experiment included two random Croatian words that served as catch trials, unlike in Experiment 1. Their purpose was to increase the Stroop interference since participants were explicitly instructed to withhold responding to the two catch words. The meanings of these catch words were never given to the participants, so it is unlikely that participants were analyzing their meanings. It is more likely that distracting words were being subjected to simple recognition (e.g., “Is this one of the two ‘words’ that I am not supposed to respond to?”). It is possible that this “catch” manipulation influences stimulus and response conflict differently, such that the exact size of each component might vary with versus without catch trials. Of course, our primary question was simply whether both types of conflict can be observed for newly learned L2 words, which was observed in the present study.

4. General Discussion

The present manuscript aimed to extend current knowledge on the source of congruency effects in a second language. Experiment 1 conceptually replicated and extended the work of Schmidt and colleagues (2018), who investigated congruency effects in a group of unbalanced Dutch-French bilinguals. They observed similarities in semantic and response processing of native and moderately fluent second languages. We aimed to test whether the same pattern could be observed in a less proficient L2. Thus, a crucial difference from the original study, adopted in the present Experiment 1, was considerably lower second language competence. Experiment 2 took this logic even further by testing whether interference effects could also be induced by novel, recently trained foreign words. To test this idea, we first administrated a relatively short and simple training which aimed to establish links between novel words and L1 colour words. In both experiments, participants performed a bilingual Stroop task. Since the present manuscript was particularly focused on the source of congruency effects, the 2-to-1 mapping procedure (De Houwer, 2003) was used. This procedure allows one to separate stimulus conflict (i.e., conflict between word and colour meanings) and response conflict (i.e., conflict between potential responses) for both first and second languages.

4.1. The present findings support L2 semantic mediation

In Experiment 1, we observed a significant stimulus conflict for L1 words (French), but response conflict failed to reach significance (see Discussion of Experiment 1). However, both stimulus and response conflict were observed for weakly-proficient L2 words (i.e., English in native French speakers). This led to the conclusion that even distracters from a weakly spoken second language can influence semantic and response processing similarly to first language words. The objective of Experiment 2 was to see whether stimulus and/or response conflict would be eliminated with recently acquired L2 words (i.e., whether the pattern of results would be the same as in Experiment 1 or not). In Experiment 2, both stimulus and response conflict effect were significant for L1 words (English). Interestingly, we observed a significant stimulus conflict for recently trained Croatian colour words (L2). These results imply that novel, recently acquired words are not only associated with the first language lexical translations, but that there is also a very early semantic mediation (Duyck & De Houwer, 2008).

Other results converge with the notion of early semantic mediation. For instance, Duyck and Brysbaert (2004) investigated a semantic number magnitude effect in forward (L1-L2) and backward (L2-L1) translation. The standard finding is faster translation of numbers that represent small quantities (e.g., two) than larger quantities (e.g., eight). In a critical experiment, participants acquired a set of pseudowords as translation equivalents of native Dutch number words. Surprisingly, a large semantic effect of number magnitude was observed in both translation directions, even though the pseudowords had been acquired only recently. Their results suggest that novel word forms are mapped into semantics even in very early stages of L2 acquisition. Additionally, research on second language acquisition has suggested that L2 words can access the corresponding semantic representation even at the early stages of language learning (e.g. Altarriba & Mathis, 1997; see Introduction). In addition, L2 words are potent enough to bias a potential response. As evidenced in

Experiment 1, an L2 word (e.g., “green” for a native French speaker) can indirectly produce the colour response linked to that colour. Thus, when “green” is presented in the colour red, both “green” and “red” are activated as potential responses that compete for selection (Klein, 1964).

On the other hand, the novel Croatian colour words adopted in Experiment 2 produced a significant response conflict effect in errors and only a marginally significant response conflict effect in response times, both of which were smaller than with L1 words. The response conflict effect was significantly reduced in Experiment 2 for both response times and errors relative to Experiment 1. However, it seems that our training procedure was efficient enough to link the novel word with first language responses (response conflict) and with the corresponding semantic representations (stimulus conflict), even with poor overall L2 proficiency. The observed hint toward response conflict (although not significant) in recently trained L2, could potentially imply the reduced size of response conflict in early stages of L2 learning. This contrasts the proposal of Kroll and Stewart (1994), who argue that, at least at early stages of vocabulary acquisition, the novel L2 words are linked to their L1 counterparts exclusively. The model assumes that the formation of a semantic link between an L2 word and its corresponding semantic representation occurs eventually, with sufficiently high language proficiency, but not so early on. If anything, our results hint at the reverse pattern. Similarly, these results are incompatible with models that assume exclusive semantic mediation early in language learning, as response conflict was observed in L2, at least for errors.

Another model that could potentially account for the present findings, is the *Bilingual Interactive Activation* model (see van Heuven et al., 1998) which assumes the existence of shared lexical systems in proficient bilinguals (unlike the Kroll and Stewart model which clearly separates L1 and L2 lexicons). This language-nonselective access results in the automatic co-activation of words from both lexicons produced by a printed word stimulus. All activated orthographic representations (even those from the irrelevant language) compete for word identification, which slows down word recognition. Top-down inhibitory control from language nodes limits the influence of such cross-language interference. However, neither the BIA model nor its extension the BIA+ model make predictions for less proficient bilinguals (i.e., with less developed L2 lexicons), which is crucial for the present manuscript, and thus it is limited in the interpretation of the present findings.

Another alternative is a model that combines the main features of the BIA model with a developmental aspect of second language acquisition, argued by Kroll and Stewart (1994), and it is known as the developmental BIA-model (BIA-d; Grainger et al., 2010). This model assumes developmental changes in connectivity between L1 and L2 word form representations and semantics as a function of exposure to L2. For instance, a native English speaker that studies French co-activates L1 word forms “chain” and “chair”, when a to-be-learned L2 word is presented (e.g., “chaise”, French for chair). L1 word forms are mutually inhibited but connected to their semantic representations. The exposure to a novel L2 word (“chaise”) co-activates the equivalent L1 form (“chair”) and the corresponding semantic representations. A direct link between L2 word and semantics is further strengthened, which consequentially modifies the connection between L1 and L2 word forms. However, the described models do not clearly indicate what is the status of non-cognates in word recognition (i.e., words that do not share overlap in pronunciation and spelling with L1 word forms), relevant for the present manuscript.

4.2. Role of response language

The asymmetry between first and second language conflict effects might partially depend on the response language (Preston & Lambert, 1969; Tzelgov et al., 1990). In this case, French colour-word interference would increase, and English colour-word interference would decrease if native French speakers were asked to produce verbal French responses. This manuscript made use of manual responses instead of verbal responses, which is necessary for the 2-to-1 mapping procedure. Possibly different patterns of result might occur with a verbal response modality. For instance, there is evidence for substantially larger Stroop interference when verbal responses are used relative to keypress responses (Augustinova et al., 2019; Glaser & Glaser, 1989; Sharma & McKenna, 1998). This response modality effect could possibly be explained by different mechanisms that underlie manual and verbal responding (Kinoshita et al., 2017). The key assignment procedure should make it easier for participants to ignore the meaning of the presented word. Participants could simply rely on matching the ink colour with the corresponding key, without lexical mediation (as would be a case in a colour naming task). An advantage of the keypress response modality is the fact that it is not inherently related with either language, which is not the case for the verbal response modality. Possibly, other differences between languages would be seen if we had used a verbal variant. For instance, L2 words like “zelena” might interfere less than L1 words (“vert” in Experiment 1 and “green” in Experiment 2, respectively), since they were not in the response set (Risko et al., 2006). Unfortunately, the 2-to-1 mapping manipulation does not lend itself well to verbal responses, but future work with alternative methodologies might address these possibilities.

4.3. Contribution of task conflict

Some authors argue for an additional contribution of task conflict in overall Stroop tasks (Augustinova et al., 2018, 2019; Kinoshita et al., 2017). Task conflict is more relevant for comparisons between colour word trials (e.g., congruent or incongruent) and certain types of neutral stimuli. Task conflict is the observation that participants respond slower to words (e.g., colour words like “red” or neutral words like “dog”) or word-like stimuli (e.g., readable non-words like “dag”) than to non-readable stimuli, such as letter strings (e.g., “xxx”) or shapes (e.g., colour boxes). This is due to the tendency to assign attention to word reading, which is an irrelevant task, rather than to color naming, the relevant task (Kalanthroff et al., 2013; Monsell et al., 2001). Task conflict effects arise for all readable items, so it could possibly occur even for the recently acquired foreign words (e.g., “zelena”). This study used only words (i.e., no non-readable neutral stimuli), so task conflict was not relevant for the present work. However, future work might also explore whether task conflict differences exist between L1 and L2 words.

4.4. Broader implications

The present results suggest that there is a certain similarity between L1 and L2 in the way they affect semantic identification and response selection (Dyer, 1971). For instance, L2 colour words produced both stimulus and response conflict, even though there was no phonological or orthographic overlap with L1 words. Indeed, the foreign words we used were quite dissimilar from their L1 equivalents (e.g., “vert”-“green” in Experiment 1; “blue”-“plava” in Experiment 2). That is, L2 words that look quite different from their L1 equivalents are potent enough to interfere with semantic and response processing. We used a “catch” trial manipulation in Experiment 2 to increase the size of the congruency effect in order to better dissociate stimulus and response conflict for the newly-trained Croatian words. As one caveat, this manipulation could hypothetically influence the results in more ways than to just increase the overall congruency effect (and slowing participants down). For example, it might be possible that this manipulation influences stimulus and response conflict differently (e.g., increasing one more than the other). The L1 results do not suggest this, showing a relatively “typical” pattern of results for the 2-to-1 mapping procedure. However, future work might aim to examine whether L2 stimulus and response conflict effects are similar without a “catch” manipulation.

Other studies showed that extensive practice can result in the formation and reinforcement of contingencies between noncognates. For instance, Geukes, Gaskell, and Zwitserlood (2015) trained participants with word-word pairs, each consisting of a German colour word and a pseudoword. Certain pairs were presented more frequently than others, which allowed participants to learn an association between a pseudoword and a colour word. In the following Stroop task, these pseudowords produced an interference effect. Similarly, our study used a relatively short (15-20 minutes) and simple training procedure, which turns out to be sufficient for linking novel words with semantics. In that vein, it seems plausible that the amount of practice, rather that cognate status might account for occurrence of conceptual links between the L2 lexicon and semantics.

Further research might aim to investigate the effectiveness of different types of L2 word acquisition training and their influence on semantic and response processing. For instance, previous studies that are conceptually related to the one presented in this manuscript used other presentation modalities to build links between stimuli, such as auditory (Altarriba & Basnight-Brown, 2012; Altarriba & Mathis, 1997) or visual (Altarriba & Knickerbocker, 2011; Lotto & de Groot, 1998). Formation of links between novel L2 words and corresponding semantic representations seems possible when participants are explicitly instructed to process the to-be-learned words, as was the case in Experiment 2. It remains unclear under which conditions (e.g., amount and type of training, instruction requirements, etc.) recently acquired foreign words can affect both semantic and response processing. Taken together, the approach to studying interference effects and its components in foreign languages might be further extended to other varieties of language training procedures.

While our study certainly indicates that semantic and response influence are observed for some L2 words, one may wonder whether the present findings could be generalized to other word types, such as less frequent, emotional, or abstract words. For instance, Altarriba and Basnight-Brown (2011) trained a group of English monolinguals with a set of concrete (e.g., jewel), emotional (e.g., angry), and abstract (e.g., virtue) words in Spanish with a similar training procedure as that administered by Altarriba and Mathis (1997; see Introduction). A following Stroop colour identification task revealed that L2 emotional words were responded to faster than L2 concrete and abstract words. This contrasted the typical emotional Stroop effect (i.e., slower response times for emotion words relative to neutral ones) observed in L1, suggesting that recently learned L2 emotional stimuli do not possess sufficiently strong and rich semantic component. However, the source of interference produced by novel L2 words with high level of abstraction has not been investigated yet.

This distinction between different word types in bilinguals is one of the core assumptions of the *Distributed Feature Model*, which states that concrete and abstract words differ in the degree of the semantic overlap across languages (de Groot, 1992). For instance, word in L1 and L2 that have a large overlap in meaning (i.e., concrete words, such as “door” in English and “porte” in French) share many conceptual features. In contrast, abstract words (e.g., “advice” in English and “conseil” in French) often have language-specific meanings and share a smaller proportion of the conceptual nodes between languages. Moreover, words with a higher level of abstraction (i.e., abstract or emotion words) are deeply encoded in L1 (Pavlenko, 2009), due to strong connection between those words and the context in which they are learned. Thus, it seems plausible that the learning of L2 abstract words should be facilitated when presented within a context, either written, visual or spoken (Altarriba & Basnight-Brown, 2012; de Groot, 1992; Schwanenflugel et al., 1992). To sum up, both learning method and word type might play important roles in establishing connections between novel L2 words and their semantic representations. The colour-word Stroop procedure, of course, is inherently limited to the use of colour-related stimuli, but “word-word

Stroop" variants also exist that can be used with any word type (Glaser & Glaser, 1982, 1989; Schmidt et al., 2013).

5. Conclusion

The present findings indicate that there is a certain similarity in the way L1 and weakly spoken L2 words influence semantic and response processing. Even for very recently learned L2 words, both stimulus and response conflict are observed. Future studies that further investigate the influence of different L2 learning methods and materials may shed additional light on this interesting question of access to semantics in L2.

References

Altarriba, J., & Basnight-Brown, D. M. (2012). The acquisition of concrete, abstract, and emotion words in a second language. *International Journal of Bilingualism*, *16*(4), 446–452. https://doi.org/10.1177/1367006911429511

Altarriba, J., & Knickerbocker, H. (2011). Chapter 2. Acquiring second language vocabulary through the use of images and words. In P. Trofimovich & K. McDonough (Eds.), *Language Learning & Language Teaching* (Vol. 30, pp. 21–48). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.30.06alt

Altarriba, J., & Mathis, K. M. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language*, *36*(4), 550–568. https://doi.org/10.1006/jmla.1997.2493

Augustinova, M., Parris, B. A., & Ferrand, L. (2019). The loci of Stroop interference and facilitation effects with manual and vocal responses. *Frontiers in Psychology*, *10*, 1786. https://doi.org/10.3389/fpsyg.2019.01786

Augustinova, M., Silvert, L., Spatola, N., & Ferrand, L. (2018). Further investigation of distinct components of Stroop interference and of their reduction by short response-stimulus intervals. *Acta Psychologica*, *189*, 54–62. https://doi.org/10.1016/j.actpsy.2017.03.009

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences*, *16*(4), 240–250. https://doi.org/10.1016/j.tics.2012.03.001

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*, *8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410 de Groot, A. M. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1001–1018. https://doi.org/10.1037/0278-7393.18.5.1001

De Houwer, J. (2003). On the role of stimulus-response and stimulus-stimulus compatibility in the Stroop effect. *Memory & Cognition*, *31*(3), 353–359. https://doi.org/10.3758/BF03194393

De Houwer, J. (2004). Spatial Simon effects with nonspatial responses. *Psychonomic Bulletin & Review*, *11*(1), 49–53. https://doi.org/10.3758/BF03206459

Duyck, W., & De Houwer, J. (2008). Semantic access in second-language visual word processing: Evidence from the semantic Simon paradigm. *Psychonomic Bulletin & Review*, *15*(5), 961–966. https://doi.org/10.3758/PBR.15.5.961

Dyer, F. N. (1971). Color-naming interference in monolinguals and bilinguals. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 297–302. https://doi.org/10.1016/S0022-5371(71)80057-9

Eidels, A., Ryan, K., Williams, P., & Algom, D. (2014). Depth of processing in the Stroop task: Evidence from a novel forced-reading condition. *Experimental Psychology*, *61*(5), 385–393. https://doi.org/10.1027/1618-3169/a000259

Geukes, S., Gaskell, M. G., & Zwitserlood, P. (2015). Stroop effects from newly learned color words: Effects of memory consolidation and episodic context. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00278

Glaser, M. O., & Glaser, W. R. (1982). Time course analysis of the Stroop phenomenon. *Journal of Experimental Psychology: Human Perception and Performance, 8*(6), 875–894.

Glaser, M. O., & Glaser, W. R. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118*, 13–42.

Grainger, J., Midgley, K., & Holcomb, P. J. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In M. Kail & M. Hickmann (Eds.), *Language acquisition across linguistic and cognitive systems* (p. 19). New York: John Benjamins.

Grosjean, F. (2012). Bilingualism: A short introduction. In *The psycholinguistics of bilingualism* (pp. 5–25). Hoboken, NJ: John Wiley & Sons.

Hasshim, N., & Parris, B. A. (2014). Two-to-one color-response mapping and the presence of semantic conflict in the Stroop task. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.01157

Jongen, E. M., & Jonkman, L. M. (2008). The developmental pattern of stimulus and response interference in a color-object Stroop task: An ERP study. *BMC Neuroscience*, *9*(1), 82. https://doi.org/10.1186/1471-2202-9-82

Kalanthroff, E., Goldfarb, L., Usher, M., & Henik, A. (2013). Stop interfering: Stroop task conflict independence from informational conflict and interference. *Quarterly Journal of Experimental Psychology*, *66*(7), 1356–1367. https://doi.org/10.1080/17470218.2012.741606

Kinoshita, S., De Wit, B., & Norris, D. (2017). The magic of words reconsidered: Investigating the automaticity of reading color-neutral words in the Stroop task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(3), 369–384. https://doi.org/10.1037/xlm0000311

Klein, G. S. (1964). Semantic power measured through the interference of words with color-naming. *The American Journal of Psychology*, *77*(4), 576. https://doi.org/10.2307/1420768

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0

Liefooghe, B., Hughes, S., Schmidt, J. R., & De Houwer, J. (2020). Stroop-like effects of derived stimulus–stimulus relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(2), 327–349. https://doi.org/10.1037/xlm0000724

Lotto, L., & de Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, *48*(1), 31–69. https://doi.org/10.1111/1467-9922.00032

MacKinnon, D. P., Geiselman, R. E., & Woodward, J. A. (1985). The effects of effort on Stroop interference. *Acta Psychologica*, *58*(3), 225–235. https://doi.org/10.1016/0001-6918(85)90022-8

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967. https://doi.org/10.1044/1092-4388(2007/067)

Marsman, M., & Wagenmakers, E.-J. (2017). Bayesian benefits with JASP. *European Journal of Developmental Psychology*, *14*(5), 545–555. https://doi.org/10.1080/17405629.2016.1259614

Monsell, S., Taylor, T. J., & Murphy, K. (2001). Naming the color of a word: Is it responses or task sets that compete? *Memory & Cognition*, *29*(1), 137–151. https://doi.org/10.3758/BF03195748

Neely, J. H., & Kahan, T. A. (2001). Is semantic activation automatic? A critical re-evaluation. In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Surprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder.* (pp. 69–93). American Psychological Association. https://doi.org/10.1037/10394-005

Pavlenko, A. (Ed.). (2009). Conceptual Representation in the bilingual lexicon and second language vocabulary learning. In *The bilingual mental lexicon: Interdisciplinary approaches*. Bristol: Multilingual Matters.

Preston, M. S., & Lambert, W. E. (1969). Interlingual interference in a bilingual version of the Stroop color-word task. *Journal of Verbal Learning and Verbal Behavior*, *8*(2), 295–301. https://doi.org/10.1016/S0022-5371(69)80079-4

Risko, E. F., Schmidt, J. R., & Besner, D. (2006). Filling a gap in the semantic gradient: Color associates and response set effects in the Stroop task. *Psychonomic Bulletin & Review*, *13*(2), 310–315. https://doi.org/10.3758/BF03193849

Schmidt, J. R., & Cheesman, J. (2005). Dissociating stimulus-stimulus and response-response effects in the Stroop task. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *59*(2), 132–138. https://doi.org/10.1037/h0087468

Schmidt, J. R., Cheesman, J., & Besner, D. (2013). You can’t Stroop a lexical decision: Is semantic processing fundamentally facilitative? *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *67*(2), 130–139. https://doi.org/10.1037/a0030355

Schmidt, J. R., Crump, M. J. C., Cheesman, J., & Besner, D. (2007). Contingency learning without awareness: Evidence for implicit control. *Consciousness and Cognition*, *16*(2), 421–435. https://doi.org/10.1016/j.concog.2006.06.010

Schmidt, J. R., & De Houwer, J. (2012). Adding the goal to learn strengthens learning in an unintentional learning task. *Psychonomic Bulletin & Review*, *19*(4), 723–728. https://doi.org/10.3758/s13423-012-0255-5

Schmidt, J. R., Hartsuiker, R. J., & De Houwer, J. (2018). Interference in Dutch–French bilinguals: Stimulus and response conflict in intra- and interlingual Stroop. *Experimental Psychology*, *65*(1), 13–22. https://doi.org/10.1027/1618-3169/a000384

Schwanenflugel, P. J., Akin, C., & Luh, W.-M. (1992). Context availability and the recall of abstract and concrete words. *Memory & Cognition*, *20*(1), 96–104. https://doi.org/10.3758/BF03208259

Sharma, D., & McKenna, F. P. (1998). Differential components of the manual and vocal Stroop tasks. *Memory & Cognition*, *26*(5), 1033–1040. https://doi.org/10.3758/BF03201181

Shichel, I., & Tzelgov, J. (2018). Modulation of conflicts in the Stroop effect. *Acta Psychologica*, *189*, 93–102. https://doi.org/10.1016/j.actpsy.2017.10.007

Sholl, A., Sankaranarayanan, A., & Kroll, J. F. (1995). Transfer between picture naming and translation: A test of asymmetries in bilingual memory. *Psychological Science*, *6*(1), 45–49. https://doi.org/10.1111/j.1467-9280.1995.tb00303.x

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Stroop, J. R. (1935). Studies on interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643–661.

Talamas, A., Kroll, J. F., & Dufour, R. (1999). From form to meaning: Stages in the acquisition of second-language vocabulary. *Bilingualism: Language and Cognition*, *2*(1), 45–58. https://doi.org/10.1017/S1366728999000140

Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling Stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(5), 760–771. https://doi.org/10.1037/0278-7393.16.5.760

Valian, V. (2015). Bilingualism and cognition. *Bilingualism: Language and Cognition*, *18*(1), 3–24. https://doi.org/10.1017/S1366728914000522

van Heuven, W. J. B., Conklin, K., Coderre, E. L., Guo, T., & Dijkstra, T. (2011). The influence of cross-language similarity on within- and between-language Stroop effects in trilinguals. *Frontiers in Psychology*, *2*. https://doi.org/10.3389/fpsyg.2011.00374

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, *39*(3), 458–483. https://doi.org/10.1006/jmla.1998.2584

Appendix: Language Demographics of Experiment 1

The majority of participants rated their language dominance (67%) and language acquisition (69%) as French first and English second. Spanish was the second most prominent second language in terms of dominance (13%) and order of acquisition (7%). Other languages listed as second in order of dominance were German, Portuguese, Arabic, and Turkish. For order of acquisition, some participants rated Turkish, German, Arabic, Italian, and Portuguese as their second language. Only one participant indicated another language (Arabic) as their native language. Since this participant rated French as their dominant language and second in order of acquisition (after Arabic), the participant sufficiently fit our language dominance criteria and was not removed from the sample. We found that 9 participants misunderstood the question about percentage of language use in the recent period, with percentages summing up to, for example, 200%, instead of 100%. These answers were treated as blanks only for this particular question and were removed from the analysis.

On average, participants self-rated their English fluency at 5.50 on a 0-10 scale. The average LEXTALE_EN score was 67.01. Most of participants correctly translated “yellow” (82/85), “green” (82/85), and “pink” (78/85). Misspelled answers (“gren”, “yelow”) and blanks were treated as incorrect. Only half of the participants (42/85) correctly translated “silver”, while the most frequent incorrect answers were “argent”, “money” and “steel”. Since “argent” is frequently used as a synonym for money in French language, this translation mistake is not surprising.