Cognitive Processing (2022) 23:619–636
https://doi.org/10.1007/s10339-022-01105-1

RESEARCH ARTICLE

Interlinguistic conflict: Word–word Stroop with first and second language colour words

Iva Šaban[1] · James R. Schmidt[1]

Received: 13 September 2021 / Accepted: 30 June 2022 / Published online: 23 September 2022

Abstract
The congruency (or Stroop) effect is a standard observation of slower and less accurate colour identification to incongruent trials (e.g. "red" in green) relative to congruent trials (e.g. "red" in red). This effect has been observed in a word–word variant of the task, when both the distracter (e.g. "red") and target (e.g. "green") are colour words. The Stroop task has also been used to study the congruency effect between two languages in bilinguals. The typical finding is that the congruency effect for L1 words is larger than that for L2 words. For the first time, the present report aims to extend this finding to a word–word variant of the bilingual Stroop task. In two experiments, French monolinguals performed a bilingual word–word Stroop task in which target word language, language match, and congruency between the distracter and target were manipulated. The critical manipulation across two experiments concerned the target language. In Experiment 1, target language was manipulated between groups, with either French (L1) or English (L2) target colour words. In Experiment 2, target words from both languages were intermixed. In both experiments, the congruency effect was larger when the distracter and target were from the same language (language match) than when they were from different languages (language mismatch). Our findings suggested that this congruency effect mostly depends on the language match between the distracter and target, rather than on a target language. It also did not seem to matter whether the language-mismatching distracter was or was not a potential response alternative. Semantic activation of languages in bilinguals and its implications on target identification are discussed.

Keywords Word–word Stroop · Bilingualism · Target language · Congruency · Language match

Introduction

In the literature on bilingual cognition, much work has focused on understanding how two languages are stored in memory and how they interact (Bialystok et al. 2008; Chen and Leung 1989; de Groot 1992; Kroll and Stewart 1994; Paivio et al. 1988; Potter et al. 1984). One tool used for studying interlinguistic interactions is the Stroop task. In the monolingual variant of the Stroop task (Stroop 1935), participants are instructed to identify the colour of a printed word (e.g. "red" printed in green), while ignoring the word itself. Even though the word meaning is irrelevant for performing the task, participants tend to respond slower and less accurately on incongruent trials (i.e. where the word and ink colour mismatch, e.g. "red" printed in green) relative to congruent trials (i.e. where word and ink colour match, e.g. "red" printed in red) and neutral trials (i.e. where distracter is colour neutral, e.g. "dog" printed in red). This finding is known as the *congruency* or *Stroop effect* (Dalrymple-Alford

This work was supported by the French "Investissements d'Avenir" program, project ISITE-BFC (contract ANR15-IDEX-0003) to James R. Schmidt. R scripts and data for the reported analyses are available on the Open Science Framework (link: https://osf.io/rg5kj/).

✉ Iva Šaban
iva.saban@u-bourgogne.fr

1 LEAD-CNRS UMR 5022, Université Bourgogne Franche-Comté, Pôle AAFE, 11 Esplanade Erasme, 21000 Dijon, France

and Budayr 1966; Logan and Zbrodoff 1979; MacLeod 1991; Schmidt and Besner 2008).

Pertinent for the current experiments, the congruency effect has also been observed in the word–word version of the Stroop task, which is similar to the colour-word Stroop, except that both the target and distracter are words. On each trial, a distracter (e.g. "red") is presented before a target (e.g. "green"). Both the distracter and target are colour words, and participants are explicitly instructed to ignore the first word and respond to the second word. Similar to the colour-word variant of the task, participants are faster to identify the target colour word when it is preceded by a congruent colour word (e.g. "green"–"green") relative to those preceded by an incongruent word (e.g. "red"–"green") or a neutral word (e.g. "new"–"green"; Glaser and Glaser 1982). Responses are also slower in the incongruent condition relative to the neutral condition (Schmidt et al. 2013).

The Stroop task has been used to study congruency effects in bilinguals (Altarriba and Mathis 1997; Dyer 1971; Preston and Lambert 1969; Schmidt et al. 2018; Tzelgov et al. 1990). The Stroop effect was observed with both native language (L1) colour words and second language (L2) colour words. For example, a native English speaker who also speaks French will be impaired by both English (e.g. "red" in green) and French incongruent colour words (e.g. "rouge" in green). The standard finding is that the congruency effect is typically larger for L1 relative to L2 words (Altarriba and Mathis 1997). This implies that the native English speaker performing the colour identification task will be more impacted by English than by French incongruent stimuli.

However, this asymmetry in the magnitude of L1 and L2 congruency effect can be modulated by different factors. One of them is a response language (Preston and Lambert 1969; Tzelgov et al. 1990), which refers to the similarity between the interfering and naming language. For instance, the response language can either match the interfering distracter language (e.g. "red" in green, where the response should be "green") or mismatch (e.g. "rouge", French for red, printed in green, where the response should be "green"). In the former example, the distracter and target are from the same language, therefore producing a within-language (intralingual) congruency effect. In contrast, the presentation of a distracter and target from different languages will result in a between-language (interlingual) congruency effect. The magnitude of within- and between-language congruency has been compared across studies. The standard finding is that the magnitude of congruency effect is larger in the within-language condition (Fang et al. 1981; Kiyak 1982; MacLeod 1991). However, the magnitudes of within- and between-language congruency effects depend on different factors, such as orthographic similarity of bilinguals' languages (and related cognate status), or subjective L2 proficiency. These factors are discussed, respectively.

First, the between-language effect is modulated by the orthographic similarity of the two languages. That is, more overlap between languages leads to stronger effects in the between-language condition (Dyer 1971; Fang et al. 1981; Preston and Lambert 1969). For instance, Preston and Lambert (1969) found that between-language interference was only 68% of the within-language interference for English-Hungarian bilinguals, but 95% for French–English bilinguals. Similarly, in the case of cognates which are translation equivalents similar in spelling and/or pronunciation across languages (e.g. "blue" in English and "bleu" in French), the between-language congruency effect (e.g. a French distracter "bleu" named in English) was almost as large as the within-language congruency effect (e.g. "blue" named in English; Dyer 1971; Preston and Lambert 1969). The same applies for the combinations of languages using different scripts. In a study with Chinese–English, Spanish–English, and Japanese–English bilinguals, Fang et al. (1981) found greater within- than between-language effects. However, languages that use the same scripts (e.g. Spanish and English) produce stronger effects in the between-language condition.

Second, the magnitudes of within- and between-language congruency effects are influenced by subjective L2 proficiency (Fang et al. 1981; Mägiste 1984; Tzelgov et al. 1990). For instance, in a group of participants much more proficient in their L1 than in their L2, Tzelgov et al. (1990, Experiment 2) observed that the congruency effect produced by L1 words was relatively large (and of comparable size) in both the within-language (L1–L1) and between-language (L1–L2) conditions. The congruency effect produced by L2 words was relatively large only in the within-language (L2–L2) condition. However, in a group of balanced bilinguals, the two within-language and between-language effects were about the same size. An interaction between orthography and proficiency was also observed. For instance, Brauer (1998) conducted Stroop studies with high and low proficiency bilinguals in languages with high (German-English) and low (English–Greek, English–Chinese) overlap. He observed that low-proficiency bilinguals showed more within- than between-language congruency effect when responding in their L1, regardless of how much the languages overlapped. However, the opposite pattern occurred when responding in their L2. On the other hand, high-proficiency participants, when speaking languages with no overlap showed greater within- than between-language congruency when responding in both languages, whereas high-proficiency bilinguals of languages with high orthographical overlap showed equal amounts of within- and between-language congruency effects (Brauer 1998). These results suggest that differences in L1 and L2 lexical processing are influenced by various factors (see also Gollan et al. 2009).

As an aside, the congruency effect seems to be less present in a keypress (i.e. manual), relative to a vocal (i.e. verbal) response modality. That is, a larger congruency effect occurs when participants are required to identify the ink colour of the printed stimulus vocally (i.e. saying the colour aloud) as compared to manually (i.e. pressing a corresponding key; Augustinova et al. 2019; Glaser and Glaser 1989; Redding and Gerjets 1977; Sharma and McKenna 1998; White 1969). The present series of experiments used manual responses exclusively, so further reasoning will focus on this particular response modality. However, we will return to this point in the General Discussion.

Especially pertinent for the current research, there is another important factor that could possibly explain the asymmetry between L1 and L2 congruency. According to the *response set membership* account (Klein 1964; Risko et al. 2006; Sharma and McKenna 1998), the magnitude of the congruency effect depends on whether a distracter is an eligible response. For instance, imagine a Stroop paradigm using the target colours "red", "blue", "green", and "yellow". In an incongruent trial such as "red" followed by "green", the distracter "red" is one of the possible targets. For this reason, "red" is expected to interfere more than colour words that are not in the response set (e.g. "brown", which is not one of the potential targets). In a cross-linguistic condition, when the distracter and target belong to different languages (e.g. "rouge"–"green"), a distracter like "rouge" is not a potential response (i.e. it is not in the response set), therefore interfering less than its English equivalent "red". To sum up, according to this view, the asymmetry between within-language (e.g. "red"–"green") and between-language (e.g. "rouge"–"green") congruency effects could be due to the fact that different-language words were not potential target responses.

A word–word variant of the Stroop task is a suitable tool for investigating the source of this asymmetry and the role of the language match and response set membership on target identification. For instance, it separates the irrelevant (i.e. to-be-ignored distracter) and relevant task dimensions (i.e. to-be-named target) temporally and spatially. It should be noted that in a standard Stroop task trial, these two dimensions are displayed simultaneously. Another modification concerns displaying both components of a standard Stroop stimulus in the same modality (i.e. both the distracter and target are words). Related to that, the language match between the distracter and the target (e.g. "red"–"green" when two languages match or "rouge"–"green" when two languages mismatch) could be manipulated. As already discussed, the fact that the two words come from the same language could increase the congruency effect (see the discussion on within- vs. between-language congruency effect above). In contrast, in the colour-word Stroop, the "language" of the target stimulus (i.e. the print colour) and therefore the language match with the distracter cannot be manipulated. Moreover, a word–word Stroop task allows us to manipulate the response eligibility of a distracter word. That is, in certain conditions, a distracter could be a potential target, which is not the case in the standard Stroop task. This could again influence the magnitude of the congruency effect, with a larger effect when the distracter is a potential response than when it is not.

Our word–word manipulation helps us to distinguish the role of these two factors (i.e. language match and response set membership) and examine their contribution to the congruency effect. In Experiment 1, we used a between-subject design. All targets were either in English or in French (depending on the group assignment). However, all participants were presented both English and French distracters. As such, participants were presented on some trials with distracters that were from a different language than the target (language mismatch, e.g. French distracters in the English target condition). These distracters were not potential targets (i.e. because, in this case, the targets were English words exclusively). In Experiment 2, however, we used a within-subject design. All participants were presented with both English and French distracters and English and French targets. This is a key difference, because a distracter that does not match in language with the target (language mismatch) could still be a potential target. For instance, if the distracter "vert" (French for "green") is followed by the target "brown", there is a language mismatch, but "vert" was a possible target stimulus on other trials. This was not the case in Experiment 1, where all targets were from the same language. In other words, all distracters belong to the response set, which should result in a larger congruency effect as compared to the one observed with a between-language manipulation. In other words, if language match between the distracter and target is all that matters, then the congruency effect should be smaller in the language mismatch condition of both experiments. If response set membership matters, then the reduction of the congruency effect in the language mismatch condition should only be observed in Experiment 1.

The influence of cross-linguistic word pairs (e.g. "red" and "rouge" in a native English speaker) on target identification can be possibly explained by the number of overlapping features between the distracter and target (de Groot 1992). According to the de Groot (1992) model, illustrated in Fig. 1, bilinguals have conceptual representations for words in both L1 and L2. These representations consist of semantic features which are distributed across languages. That is, translation equivalents possess both shared and separate meaning components. More relevant for the present research is, however, the assumption that semantic representation is richer for L1 than for L2 words. This could suggest an overall larger effect for the L1 words. According to the model, L1

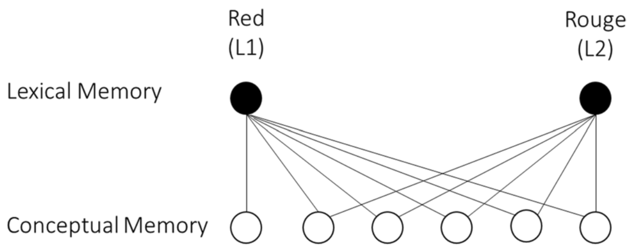

Fig. 1 Distributed conceptual representations in bilingual memory assumed by the de Groot (1992) model

words activate more semantic features than L2 words, thus producing a larger priming effect to L2 words (L1–L2) than vice versa (Schoonbaert et al. 2009). However, the congruency effect in a word–word Stroop is expected to be larger in L1 since L1 words are strongly activated by the conceptual (semantic) system (de Groot 1992; Green 1986, 1998). The incongruent colour words (e.g. "red" and "green") therefore activate a large number of overlapping semantic nodes, thus impairing a target identification. It is plausible therefore that a larger overall effect could be observed for L1 words, regardless of target language.

This manuscript aimed to examine the role of other factors that can possibly influence target colour identification in a word–word variant of the Stroop task. As briefly mentioned, one potential factor is *target language*. For instance, L1 targets are expected to be responded faster to than L2 targets. A second factor is *language match*, which refers to whether the distracter language matches the target language. As already discussed, trials in which the distracter and target language mismatch should be responded to faster relative to trials in which distracter and target belong to the same language. Third, *response set membership* might influence the congruency effect, with smaller effects on language-mismatch trials, but only if the distracter is not a potential response (i.e. as in Experiment 1, but not in Experiment 2). More trivially, a fourth factor is *congruency*, which refers to the match or mismatch in the colour concepts activated by the distracter and target. In line with previous reasoning, responses on congruent trials (i.e. when the distracter and target refer to the same colour) are assumed to be faster than those on incongruent trials (i.e. when the distracter and target refer to different colours).

To sum up, the present manuscript aimed to identify the factors underlying the L1–L2 asymmetry using the word–word variant of the Stroop task, which has not been done previously. In this variant, both the distracter and target are words. This is not the case in a standard Stroop task where to-be-ignored distracter is a word and the to-be-attended target is colour (i.e. language-neutral). This important feature of the word–word Stroop task allowed us to manipulate two factors that could account for the L1–L2 asymmetry: language match and response set membership. In Experiment 1, targets were either French (L1) or English (L2) words presented in between-subject design. Based on the previous findings from the bilingual Stroop literature, the congruency should be larger when the two words come from the same language (within-language condition) than when they are from different languages (between-language condition). That is, the congruency effect is expected to be larger on French–French or English–English trials (within-language) relative to French–English or English–French (between-language) trials. However, this asymmetry could be due to the fact that different-language words are not potential targets (e.g. "marron"–"green" in the English target condition, where "marron" was not in the response set). As already discussed, the response-set membership account predicts that the congruency effect should be smaller when distracter is not a potential response. In this case, different-language words are expected to interfere less than same-language words, but only when the different-language words are not in the response set (e.g. "marron"–"green", when "marron" is not in the response set). In Experiment 2, all distracters were presented as possible targets. French and English targets occurred interchangeably in the within-subject design, and both language words are considered as possible targets. If response-set membership is the key factor, then this manipulation should not reveal a reduced congruency effect for different-language words. Indeed, congruency effects for different-language words should be comparable to those of same-language words, or at least larger than the congruency effects for different-language words in Experiment 1, since all distracters are potential targets in Experiment 2. For instance, for the stimulus "marron"–"green", the distracter "marron" is from a different language but could be a potential target. That is, "marron" should produce a congruency effect of comparable magnitude as its same-language equivalent "brown". However, if the language match between distracter and target matters, then the "marron"–"green" trial should produce a much smaller effect than a "brown"–"green" trial or a "marron"–"vert" trial in Experiment 2, just as in Experiment 1. Though we deemed it less likely, it is also possible that the congruency effect is simply larger for L1 than for L2, and neither language match or response-set membership are relevant factors. In this case, we would anticipate larger overall French congruency effects and no effects of language match in either experiment.

Experiment 1

Experiment 1 aimed to investigate the way target language, language match, and congruency between the distracter and target influence colour word identification. Our participants performed a word–word variant of the Stroop task, in which a colour word distracter preceded a to-be-identified colour word target. A critical manipulation concerned the target language, that is, participants were randomly assigned either to the English or French target condition. In other words, participants indicated the target colour identity of English words ("green", "brown", "pink", or "white") in the English-target condition and the target colour identity of French words ("vert", "marron", "rose", or "blanc", respectively) in the French-target condition. In both groups, they needed to ignore the distracter that was presented either in the matching (i.e. English distracter-English target or French distracter-French target) or mismatching language (i.e. English distracter-French target or French distracter-English target).

Method

Participants

A total of 81 University of Burgundy undergraduates (70 women, 10 men, 1 unknown) participated in the study ($MEAN_{age}$ = 19.51, SE = 0.29). They were recruited on social networks or university studying platforms and received course credit for their participation. The only requirement for participation was to be a native French speaker. Language questionnaires (see *Results* section) were used to confirm the fit of participants with this criterion. Participants performed a single experimental session which lasted around 25–30 min.

Apparatus and materials

The experiment was run online. Stimulus presentation and response collection were controlled by Psytoolkit software (Stoet 2010, 2017). Prior to the experimental portion, participants filled out a series of questions from the French version of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al. 2007). The first three questions were retained, which asked participants to list their languages in order of dominance and in order of acquisition. Also retained from the LEAP-Q was a box asking for the age that the participants began acquiring French, became fluent in French, began learning to read French, and became fluent in reading French. As an addition to this questionnaire, participants were asked to indicate their age, sex, and native language. They also self-rated their English competence on a 1–5 scale (1 = almost none; 5 = perfect) and indicated the number of years they had studied English in school. These language metrics scores were correlated with the observed congruency effects. Finally, to assure that participants were familiar with the English colour words used in the experiment ("green", "brown", "pink", and "white"), they were asked to give their French translations.

This questionnaire portion of the experiment was followed by the English version of LexTALE vocabulary test (Lemhöfer and Broersma 2012) with French instructions. In this test, participants were presented with a list of 60 English-looking words, only about 2/3 of which were actual English words (e.g. "scholar"), whereas the remaining 1/3 were not (e.g. "kilp"). The participants were instructed to select the words that they are fairly certain are actual English words by pressing the "F" key. Otherwise, they were to press the "J" key to indicate that they did not think it was an existing English word. Correct responses were awarded with one point and incorrect "false alarms" were penalized by two points.

Design

During the main part of the experiment, participants were presented with French and English colour words. French/English colour word equivalents were "vert"/"green", "marron"/"brown", "rose"/"pink", and "blanc"/"white". The presentation of these colour words varied across three factors. The *target language* factor was manipulated between groups. Participants were randomly assigned to either the "French target" or "English target" condition. The two within-group factors were *language match* (with 2 levels: *same*, in which the distracter and target are from the same language; and *different*, in which the distracter and target are from different languages) and *congruency* (with 2 levels: *congruent*, in which the distracter and target refer to the same colour; and *incongruent*, in which the distracter and target refer to different colours).

The experimental portion of the study consisted of one practice block and four 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 and was followed by either a French or English target colour word, depending on the condition. There were also 512 experimental trials with 128 trials per block. The 32 possible trials (i.e. 8 distracters × 4 targets) were presented 4 times within each block, and each set of 32 trials was randomized without replacement. In the "French target" condition, the target stimuli were always French colour words, which could be preceded by either a French or English colour word. Similarly, in the "English target" condition, the target stimuli were always English colour words, preceded by either a French or English colour word.

Procedure

After completing the survey and LexTALE (see above), the main part of the experiment began. Each trial started with the fixation ("+") presented in the centre of the screen for

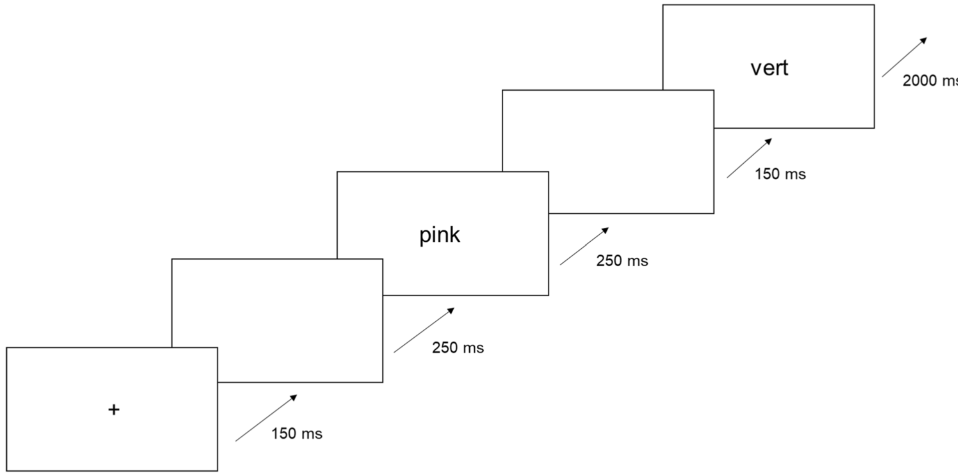

Fig. 2 An example experimental trial with corresponding timings

150 ms. This was followed by a blank screen for 250 ms. The prime stimuli (either "xxxx" in the practice block or the French/English colour word in the experimental block) was then presented in the centre of the screen for 250 ms. This was replaced by a blank screen for 250 ms. Finally, the target colour word appeared on 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 500 ms before the next trial. The procedure is visualised in Fig. 2. For each participant, regardless of the condition they were assigned to, the four colours had fixed key mapping: green ("c"), brown ("v"), pink ("b"), and white ("n").[1]

Results

Language demographics

All participants were native French speakers (100%). For almost all participants, French was the first language in order of dominance (93.83%) and in order of acquisition (96.3%). Participants mostly indicated English (80.25%), Spanish (9.88%), and French (2.47%) as a second language in order of dominance. Other languages such as German, Creole, and Turkish, as well as "unknown" cells were represented in low percentages (in total 7.4%). As a second language in order of acquisition, participants indicated English (80.25%), Spanish (4.94%), German (2.47%), French (2.47%), and Italian (2.47%). Other languages (Creole, Turkish, Arabic, Vietnamese, and Portuguese) and "no answer" cells accounted for 7.4% of total responses. Participants are highly exposed to French: 89% of them rated the amount of daily exposure between 80 and 100% of time. Mean age (in years) of French speaking acquisition was 1.61 (SE = 0.17), and fluent speaking was 3.8 (SE = 0.23). The participants started reading on average at age (in years) of 5.32 (SE = 0.13), while level of fluent reading they achieved at age of 6.81 (SE = 0.19).

Participants self-rated their English proficiency moderately (MEAN = 3.01, SE = 0.09) on 1–5 scale. All of them had studied English in school (MEAN = 9.81, SE = 0.25). Performance on the objective English vocabulary test (LexTALE) was average (MEAN = 68.54, SE = 1.12). Participants were familiar with the English colour words used in the Stroop task. They were highly accurate in translating *pink* (100%), *green* (98.67%), *brown* (98.67%), and *white* (96.3%).

Stroop task response times

The data were analysed in a three-way mixed analysis of variance (ANOVA) with repeated measures on the following factors: target language (French vs. English), language match (same vs. different) and congruency (congruent vs. incongruent). Target language was manipulated at a between-subject level, and the remaining factors (i.e. language match and congruency) at a within-subject level. Only correct responses were analysed. In the French target condition, 8.65% of the trials were excluded (1% of time-out trials and 7.65% of incorrect trials). In the English target condition, we excluded 7.17% of trials from the analysis (0.88% of time-out and 6.29% of incorrect trials). The mean RT data are presented in Fig. 3.

There was a significant main effect of language match, $F(1,79) = 24.118$, $p < 0.001$, $\eta^2 p = 0.234$, MSE = 1037.953, $BF_{10} = 27.771$, indicating faster responses when the distracter and target were from the same language relative to when they were from different languages, $t(80) = 4.92$,

[1] No specific instructions on hand/finger placement were given. However, typically participants spontaneously use the middle and index fingers of the left (for "c" and "v" keys) and right (for "b" and "n" keys) hands.

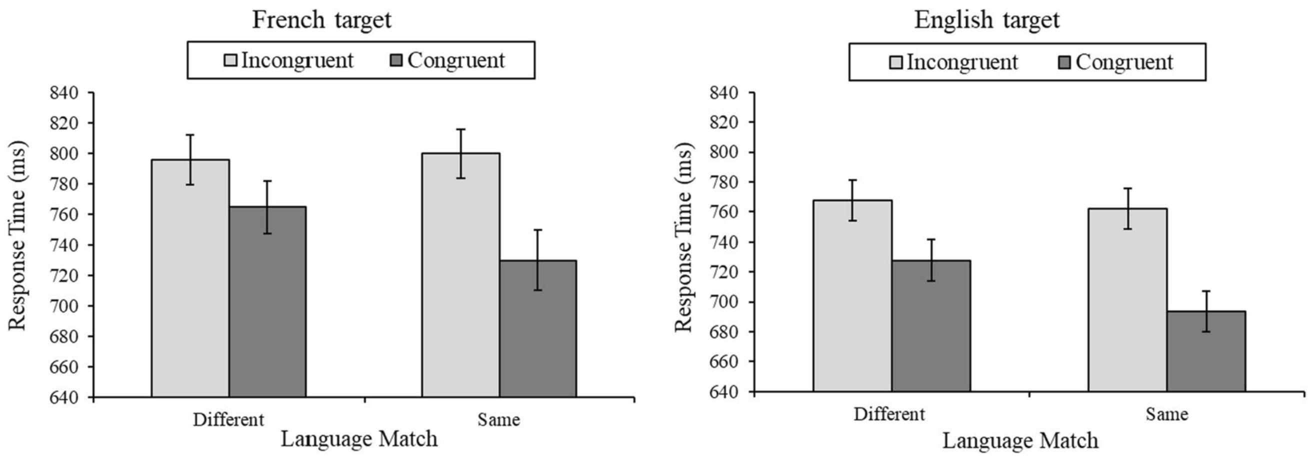

Fig. 3 Mean response times with standard errors for French and English target language condition

Table 1 Mean response times and standard errors (in brackets) for each type of trials

	French target condition		English target condition	
	Different language	Same language	Different language	Same language
Incongruent	795.7 (16.5)	799.6 (16.2)	767.5 (13.6)	762.3 (13.5)
Congruent	764.5 (17.3)	729.7 (19.7)	727.8 (13.9)	693.5 (13.8)

$p < 0.001$, $MEAN_{diff} = -17.6$, $SE_{diff} = 3.57$, Cohen's $d = -0.547$, $BF_{10} > 100$. We also observed a main effect of congruency, $F(1,79) = 141.355$, $p < 0.001$, $\eta^2 p = 0.641$, $MSE = 1570.93$, $BF_{10} > 100$, indicating faster responses on congruent as compared to incongruent trials, $t(80) = 11.9$, $p < 0.001$, $MEAN_{diff} = -52.3$, $SE_{diff} = 4.38$, Cohen's $d = -1.33$, $BF_{10} > 100$. Surprisingly, there was no main effect of target language, $F(1,79) = 2.63$, $p > 0.05$, $\eta^2 p = 0.032$, $MSE = 36{,}841.118$, $BF_{10} = 0.918$, $BF_{01} = 1.089$, indicating no overall difference in response speed between French and English target words.

There was a statistically significant two-way interaction between language match and congruency, $F(1,79) = 26.990$, $p < 0.001$, $\eta^2 p = 0.255$, $MSE = 857.482$, $BF_{10} > 100$, indicating that the congruency effect was larger in the same language condition than in the different language condition. The congruency effect was significant in both the different language condition, $t(80) = 8.39$, $p < 0.001$, $MEAN_{diff} = 35.4$, $SE_{diff} = 4.22$, Cohen's $d = 0.933$, $BF_{10} > 100$, and in the same language condition, $t(80) = 10.7$, $p < 0.001$, $MEAN_{diff} = 69.267$, $SE_{diff} = 6.45$, Cohen's $d = 1.19$, $BF_{10} > 100$.

The three-way interaction between target language, language match, and congruency was not significant, $F(1,79) = 0.547$, $p > 0.05$, $\eta^2 p = 0.007$, $MSE = 857.482$, $BF_{10} = 0.233$, $BF_{01} = 4.292$. Mean response times and standard errors for all combinations of these three factors are displayed in Table 1. Neither the interaction between target language and language match, $F(1,79) = 0.368$, $p > 0.05$, $\eta^2 p = 0.005$, $MSE = 1037.954$, $BF_{10} = 0.158$, $BF_{01} = 6.329$, nor the interaction between target language and congruency, $F(1,79) = 0.179$, $p > 0.05$, $\eta^2 p = 0.002$, $MSE = 1570.93$, $BF_{10} = 0.14$, $BF_{01} = 7.143$, were significant. As such, both groups were influenced by language match, but did not seem to differ otherwise.

Stroop task percentage error

The percentage error data are presented in Fig. 4. We observed a significant main effect of language match, $F(1,79) = 5.828$, $p < 0.05$, $\eta^2 p = 0.069$, $MSE = 6.533$, $BF_{10} = 0.928$, $BF_{01} = 1.077$, indicating less accurate responding when the distracter and target were from the same language as compared to when they were from different languages, $t(80) = 2.35$, $p < 0.05$, $MEAN_{diff} = 0.679$, $SE_{diff} = 0.289$, Cohen's $d = 0.261$, $BF_{10} = 1.64$. However, there was no main effect of target language, $F(1,79) = 2.28$, $p > 0.05$, $\eta^2 p = 0.028$, $MSE = 119.672$, $BF_{10} = 0.878$, $BF_{01} = 1.139$ or even congruency, $F(1,79) = 0.351$, $p > 0.05$, $\eta^2 p = 0.004$, $MSE = 11.553$, $BF_{10} = 0.158$, $BF_{01} = 6.329$.

The three-way interaction between target language, language match, and congruency was not significant, $F(1,79) = 0.08$, $p > 0.05$, $\eta^2 p = 0.001$, $MSE = 7.684$, $BF_{10} = 0.316$, $BF_{01} = 3.164$. Mean percentage errors and standard errors for all combinations of these factors are displayed in Table 2. The two-way interaction between language match and congruency failed to reach significance, $F(1,79) = 0.042$, $p > 0.05$, $\eta^2 p = 0.001$, $MSE = 7.684$,

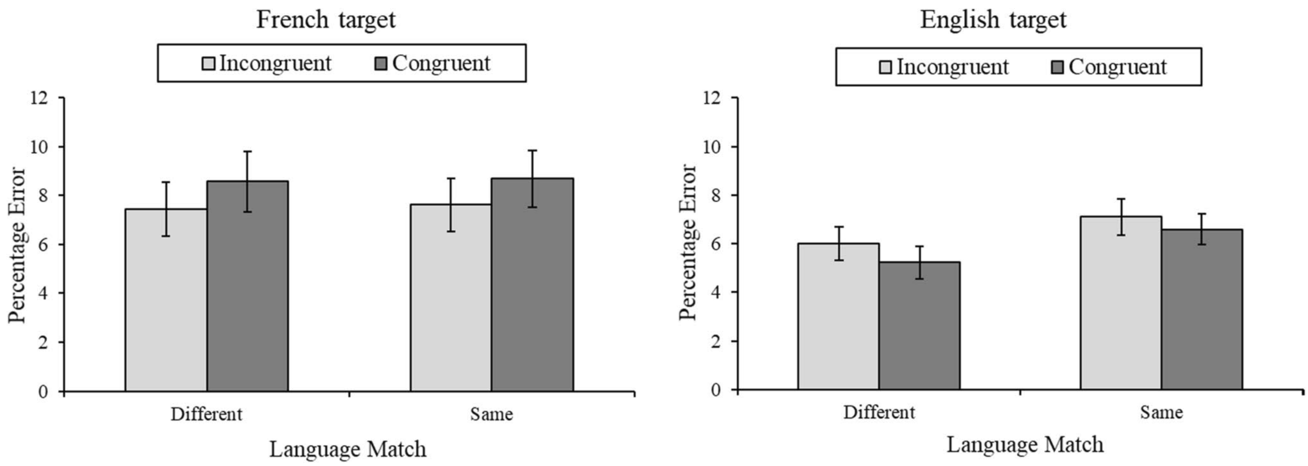

Fig. 4 Mean percentage errors with standard errors for French and English target language condition

Table 2 Mean percentage errors and standard errors (in brackets) for each type of trials

	French target condition		English target condition	
	Different language	Same language	Different language	Same language
Incongruent	7.4 (1.1)	7.6 (1.1)	6.0 (.7)	7.1 (.7)
Congruent	8.6 (1.2)	8.7 (1.1)	5.2 (.6)	6.6 (.6)

$BF_{10}=0.176$, $BF_{01}=5.682$, indicating that the relationship between congruency effect and the accuracy did not depend on the match between the language of distracter and target.

The two-way interaction between target language and language match was marginally significant, $F(1,79)=3.606$, $p=0.061$, $\eta^2p=0.044$, $MSE=6.533$, $BF_{10}=0.533$, $BF_{01}=1.876$. In the French target condition, there was no significant difference in error rates between same and different language match, $t(40)=0.338$, $p>0.05$, $MEAN_{diff}=0.146$, $SE_{diff}=0.433$, Cohen's $d=0.053$, $BF_{10}=0.178$, $BF_{01}=5.618$. In the English target condition, participants had significantly lower error rates when prime and target were from different languages than when they were from the same language, $t(39)=3.36$, $p<0.01$, $MEAN_{diff}=1.23$, $SE_{diff}=0.365$, Cohen's $d=0.530$, $BF_{10}=18.3$.

The two-way interaction between target language and congruency was significant, $F(1,79)=5.352$, $p<0.05$, $\eta^2p=0.063$, $MSE=11.553$, $BF_{10}=4.276$. In the French target condition, participants made marginally more errors to congruent relative to incongruent trials, $t(40)=2.08$, $p<0.05$, $MEAN_{diff}=1.10$, $SE_{diff}=0.528$, Cohen's $d=0.325$, $BF_{10}=1.17$. In English target condition, there was no significant difference in percentage error between congruent and incongruent trials, $t(39)=1.209$, $p>0.05$, $MEAN_{diff}=0.65$, $SE_{diff}=0.537$, Cohen's $d=-0.19$, $BF_{10}=0.333$, $BF_{01}=3.003$.

Correlations

As an additional analysis, we tested the level to which language demographic data collected in the initial portion of the study correlate with the congruency effects measured in the experimental portion. The language demographic data were collected through 1) the LexTALE English vocabulary test considered as an objective measure of L2 proficiency, and 2) a set of questions taken from the LEAP-Q, which asked for participants' self-ratings (e.g. English level, French exposure, etc.) and estimations (e.g. age of French acquisition, fluent reading, etc.). Thus, we tested the correlations between language demographic data and congruency effect measures in the different experimental conditions. We note that these demographic data were primarily collected for the selection criteria of the experiment (i.e. to assure that our participants were dominant L1 speakers), but we present the following correlations for information purposes. The non-parametric rank-based Spearman's correlation coefficients are presented in Table 3. The two largest correlation coefficients (significant at 0.01 level) were found between error performance measures (when French target was preceded by incongruent French word, e.g. "vert"–"marron", or incongruent English word, e.g. "green"–"marron") and LexTALE score. However, none of the correlations reached significance at 0.05 level after applying a Holm-Bonferroni correction for multiple comparisons, which suggests that these correlations should be interpreted with caution.

Table 3 Correlations between performance (response times and errors) and language measures

	French target condition								English target condition							
	Different language				Same language				Different language				Same language			
	Incongruent		Congruent		Incongruent		Congruent		Incongruent		Congruent		Incongruent		Congruent	
	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR
LexTALE	−.064	−**.481**	.024	−.299	.054	−**.472**	.115	−*.381*	−.150	−.051	−.131	−.154	−.177	−.089	−.065	−.060
English Level	.148	−.077	.173	.140	.177	−.060	.197	−.033	.057	.198	−.047	.228	−.005	.248	.053	−.032
Years English	.046	.194	.042	.249	.001	.095	.025	−.061	−.225	.098	−.163	.005	−.187	.072	−.206	−.216
% French exposure	−.070	−.111	−.119	−.004	−.111	−.045	−.119	−.136	−.166	−.232	−*.319*	−.130	−.223	−.028	−.233	.003
FRENCH																
Acquisition	−.078	.108	−.069	.099	−.128	.275	.023	.072	.059	.004	−.052	.046	−.019	.068	−.044	.274
Fluent	.056	−.106	.056	.194	−.005	.112	.091	.106	.145	−.116	.134	.026	.002	−.110	.037	.036
Reading	.019	−.124	−.016	.078	−.003	−.068	.032	.136	−.066	.037	−.025	−.010	.034	−.055	−.092	−.100
Fluent Reading	.209	.071	.189	.280	.218	.112	*.332*	.159	.006	.000	.123	−.076	.039	−.055	−.045	−.100

Italic = $p < .05$, Bold = $p < .01$. No tests were significant after a Holm–Bonferroni correction

Discussion

Experiment 1 showed no difference in target identification speed on French (L1) and English (L2) target words. The target language does not seem to matter in colour identification. Further, the congruency effect was not robustly larger for L1 than for L2, consistent with the idea discussed in the Introduction that the presence of an asymmetry between L1 and L2 depends on the response language. However, language match between the distracter and target had a robust influence on behaviour. That is, the congruency effect was larger in the same-language condition (i.e. when the distracter and target belonged to the same language) than in the different-language condition (i.e. when the distracter and target belonged to different languages). This confirms previous findings of larger within-language relative to between-language congruency effects (Fang et al. 1981; Kiyak 1982; MacLeod 1991; Preston and Lambert 1969). This finding could also be considered consistent with both the language match and the response set membership accounts discussed in the Introduction (which will be dissociated in Experiment 2). In both the same-language and different-language conditions, congruent trials are responded to faster than incongruent trials. This could be explained by the strong overlap in semantic nodes activated by translation equivalents (e.g. "green-vert" or "vert-green") in different-language condition (Costa et al. 1999; Costa and Caramazza 1999; de Groot 1992). The same pattern observed in the same-language condition confirms the findings from the lexical decision literature, suggesting the faster identification of words preceded by physically identical words (e.g. "green-green" or "vert-vert") relative to different word (e.g. "marron-green"; Jacobs et al. 1995; La Heij et al. 1985; Perea et al. 2014).

Experiment 2

Experiment 2 conceptually replicates Experiment 1 with target language being manipulated as within-subject factor. That is, all participants saw both French and English words as distracters (as in Experiment 1), but also both French and English words as targets (unlike Experiment 1). The logic of this experiment is simple. If the reason why between-language congruency is smaller than within-language congruency effect is due to the fact that different-language words were not potential target responses (i.e. they are out of the response set), then the same asymmetry should no longer be observed if both language words can also be targets. For example, for a trial like "vert"–"brown", "vert" was not a potential target in Experiment 1. In Experiment 2, "vert" might be from a different language than the target ("brown"), but "vert" can be a potential target. Thus, congruency effect should be similar (or at least much larger than in Experiment

Table 4 Mean language scores with standard errors

	Acquisition		Fluent		Reading		Fluent Read	
French (L1)	.66	(.15)	2.90	(.30)	5.31	(.21)	6.57	(.31)
English (L2)	10.28	(.51)	18.70	(.80)	12.20	(.46)	18.19	(.89)

1). In contrast, if it is the matching of the stimulus languages that matters, then a trial like "vert"–"brown" should produce weaker congruency effect than a trial like "green"–"brown". As such, results should be similar or identical to those in Experiment 1.

Method

Participants

A total of 35 participants (27 women and 8 men) took part in Experiment 2 (MEAN = 30.14, SE = 1.34). None of them participated in Experiment 1. They were all volunteers, recruited via social networks and the *Info du Risc* platform (a French academic diffusion list). The inclusion criteria and duration of the experiment were identical to Experiment 1.

Apparatus, materials, design, and procedure

The experiment was identical to Experiment 1 in all respects with a single exception. All factors were manipulated in the within-subject manner. The mixed-target language condition therefore consisted of 64 possible trials presented in random order, twice within each experimental block. In other words, participants saw all the trials from both the French targets and English targets conditions of Experiment 1, intermixed together. The LEAP-Q questions for English were also added (omitted by accident in Experiment 1).

Results

Language demographics

All participants were native French speakers (100%). For almost all of them, French was the first language in order of dominance (97.14%) and in order of acquisition (100%). The vast majority of participants indicated English (88.57%) as their second language in order of dominance, followed by Italian (5.71%), Spanish (2.86%), and Creole (2.86%). The most frequent second languages in order of acquisition were English (74.28%), Spanish (8.57%), and German (5.71%). Other responses were Italian, Creole, and Japanese. Participants are highly exposed to French in their everyday lives; 77% of them rated the amount of daily exposure between 81 and 100% of time and 14% between 61 and 80% of time. Mean French (L1) and English (L2) language metric scores are presented in Table 4.

Participants rated their English proficiency as average (MEAN = 3.31, SE = 0.16) on a 1–5 scale. All of them had studied English in school (MEAN = 9.00 years, SE = 0.47). Their performance on the LexTALE vocabulary test was relatively good (MEAN = 76.63, SE = 1.8). Participants were mostly able to correctly translate the given English colour words. The accuracy per word was high; *green* (100%), *pink* (100%), *white* (100%), and *brown* (88.57%).

Stroop task response times

The data were analysed in a three-way analysis of variance with repeated measures on the following factors: target language (French vs. English), language match (different vs. same), and congruency (incongruent vs. congruent). All the factors were manipulated at the within-subject level. Only correct responses were analysed. A total of 5.46% of incorrect trials and 1.86% of time-out trials were removed. The mean RT data are presented in Fig. 5.

There was a main effect of target language, $F(1,34) = 5.431$, $p = 0.03$, $\eta^2 p = 0.138$, MSE = 2106.934, $BF_{10} = 3.071$, indicating faster responses to French relative to English target words, $t(34) = 2.33$, $p < 0.05$, $MEAN_{diff} = 12.8$, $SE_{diff} = 5.49$, Cohen's $d = 0.394$, $BF_{10} = 1.93$. We also observed a significant main effect of language match, $F(1,34) = 23.343$, $p < 0.001$, $\eta^2 p = 0.407$, MSE = 1209.013, $BF_{10} > 100$. Participants responded significantly faster when the distracter and target were from the same language relative to when they were from different languages, $t(34) = 4.83$, $p < 0.001$, $MEAN_{diff} = -20.1$, $SE_{diff} = 4.16$, Cohen's $d = -0.817$, $BF_{10} > 100$. Finally, there was the main effect of congruency, $F(1,34) = 11.418$, $p < 0.01$, $\eta^2 p = 0.251$, MSE = 2486.531, $BF_{10} > 100$, indicating faster responses on congruent relative to incongruent trials, $t(34) = 3.38$, $p < 0.01$, $MEAN_{diff} = -20.1$, $SE_{diff} = 5.96$, Cohen's $d = -0.571$, $BF_{10} = 18.5$.

The only significant interaction was the one between language match and congruency, $F(1,34) = 12.714$, $p = 0.001$, $\eta^2 p = 0.272$, MSE = 1458.088, $BF_{10} = 66.653$, indicating that the congruency effect (incongruent–congruent) was more pronounced in same language condition. Indeed, the congruency effect was significant in the same language condition, $t(34) = 4.95$, $p < 0.001$, $MEAN_{diff} = 36.4$, $SE_{diff} = 7.36$, Cohen's $d = 0.836$, $BF_{10} > 100$, but not in the different language condition, $t(34) = 0.505$, $p > 0.05$, $MEAN_{diff} = 3.87$, $SE_{diff} = 7.95$, Cohen's $d = 0.085$, $BF_{10} = 0.204$, $BF_{01} = 4.902$.

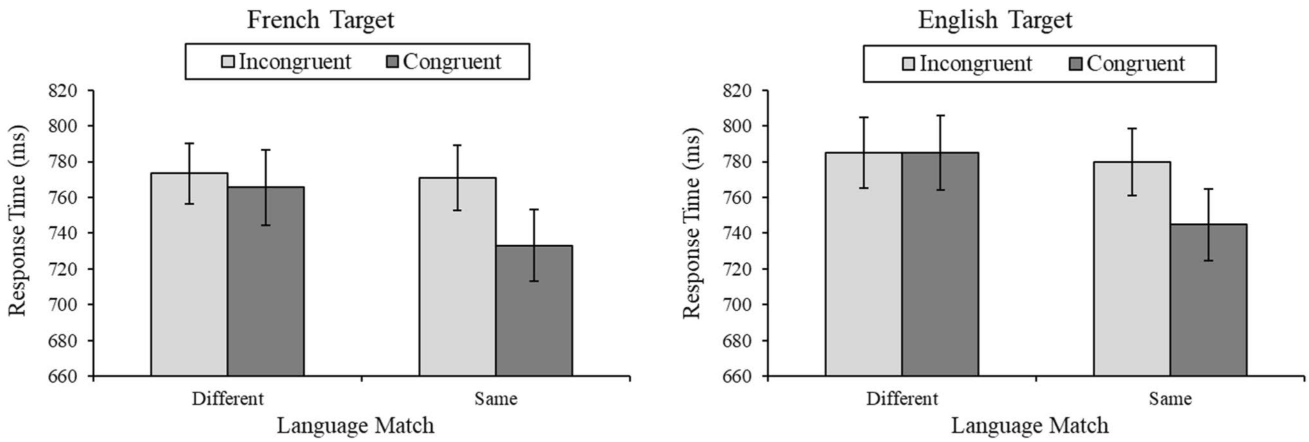

Fig. 5 Mean response times with standard errors for French and English target language condition

Table 5 Mean response times and standard errors (in brackets) for each type of trials

	French target condition		English target condition	
	Different language	Same language	Different language	Same language
Incongruent	773.4 (16.9)	771.0 (18.1)	784.8 (19.7)	779.6 (18.7)
Congruent	765.4 (21.0)	733.1 (20.1)	785.0 (20.7)	744.7 (20.0)

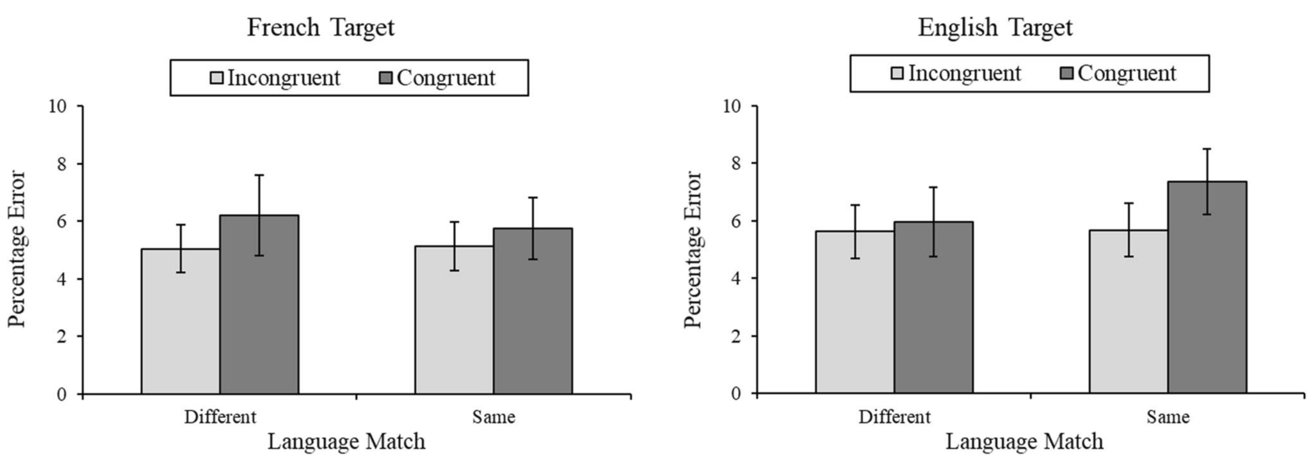

Fig. 6 Mean percentage errors with standard errors for French and English target language condition

The three-way interaction between target language, language match and congruency was not significant, $F(1,34) = 0.142$, $p > 0.05$, $\eta^2 p = 0.004$, $MSE = 832.374$, $BF_{10} = 0.279$, $BF_{01} = 3.584$. Mean response times and standard errors for all combinations of these three factors are displayed in Table 5. There was no significant interaction between target language and language match, $F(1,34) = 0.488$, $p > 0.05$, $\eta^2 p = 0.014$, $MSE = 1044.181$, $BF_{10} = 0.209$, $BF_{01} = 4.785$, or between target language and congruency, $F(1,34) = 0.622$, $p > 0.05$, $\eta^2 p = 0.018$, $MSE = 865.198$, $BF_{10} = 0.218$, $BF_{01} = 4.587$. Thus, again, language match seemed to be the only relevant variable affecting performance.

Stroop task percentage error

The mean percentage error data are presented in Fig. 6. The only significant effect in the percentage error analyses was congruency, $F(1,34) = 4.819$, $p < 0.05$, $\eta^2 p = 0.124$, $MSE = 13$, $BF_{10} = 1.052$. Surprisingly, congruent trials had higher percentage error (i.e. participants were less accurate) than in incongruent trials, $t(34) = 2.20$, $p < 0.05$,

Table 6 Mean percentage errors and standard errors (in brackets) for each type of trials

	French target condition		English target condition	
	Different language	Same language	Different language	Same language
Incongruent	5.0 (.8)	5.1 (.8)	5.6 (.9)	5.7 (.9)
Congruent	6.2 (1.4)	5.7 (1.1)	5.9 (1.2)	7.3 (1.1)

$MEAN_{diff} = 0.947$, $SE_{diff} = 0.431$, $Cohen's\ d = 0.371$, $BF_{10} = 1.51$. There were no significant main effects of target language, $F(1,34) = 3.291$, $p > 0.05$, $\eta^2 p = 0.088$, $MSE = 7.986$, $BF_{10} = 0.308$, $BF_{01} = 3.247$, or language match, $F(1,34) = 0.315$, $p > 0.05$, $\eta^2 p = 0.009$, $MSE = 16.514$, $BF_{10} = 0.156$, $BF_{01} = 6.41$.

The three-way interaction between target language, language match, and congruency was not significant, $F(1,34) = 0.949$, $p > 0.05$, $\eta^2 p = 0.027$, $MSE = 16.343$, $BF_{10} = 0.355$, $BF_{01} = 2.817$. Mean percentage errors and standard errors for all combinations of these factors are displayed in Table 6. As in the mean RT data, the interaction between target language and language match was not significant, $F(1,34) = 1.022$, $p > 0.05$, $\eta^2 p = 0.029$, $MSE = 14.233$, $BF_{10} = 0.283$, $BF_{01} = 3.533$. There was no significant interaction between target language and congruency, $F(1,34) = 0.017$, $p > 0.05$, $\eta^2 p = 0.000$, $MSE = 12.183$, $BF_{10} = 0.189$, $BF_{01} = 5.291$. Similarly, the interaction between language match and congruency was not significant, $F(1,34) = 0.145$, $p > 0.05$, $\eta^2 p = 0.004$, $MSE = 19.12$, $BF_{10} = 0.207$, $BF_{01} = 4.831$.

Correlations

As in Experiment 1, we tested the correlations of language demographic data obtained through the LexTALE test and a set of questions from the LEAP-Q questionnaire with the congruency effects measured in the experimental portion of the study. The nonparametric rank-based Spearman's correlation coefficients are presented in Table 7. As can be observed, there are some performance measures (response time or error) that correlated with English level, and percentage of French and English language use. Response time (but not error) measures correlated significantly with age of French (speaking) acquisition, fluency, age of reading acquisition and age of fluent reading. No significant correlations were observed for English LEAP-Q age measures. Some correlations were significant at $a = 0.001$ level (for instance, the age of reading acquisition and fluent reading in French correlated with response times when French target word is preceded by incongruent French distracter, e.g. "vert"–"marron"). However, these correlations should be interpreted with caution, since after applying a Holm–Bonferroni correction for multiple comparisons, none of them reached significance at the 0.05 level.

Discussion

Experiment 2 is a conceptual replication of Experiment 1 with an intermixed presentation of both L1 and L2 target words for all participants. Both L1 and L2 words were therefore presented as potential distracters and targets, which made them a part of the response set (Klein 1964; Risko et al. 2006). The main effect of target language was observed, with faster responses on French (L1) than English (L2) target words. It is plausible that L1 words are strongly activated by the semantic system, which facilitates responses to L1 targets (Green 1986, 1998). More importantly, the interaction between language match and congruency was again significant. However, no congruency effect occurred in the different-language condition. Once again, the within-language congruency is much larger than between-language congruency effect. This contradicts the assumption of the response set membership account, which assumes that both language distracters should interfere equally (or, at minimum, that different language distracters should produce are notably larger congruency effect than that observed in Experiment 1), since all distracters are potential targets. Interestingly, a language match effect is still present even when all distracters are potential targets. This suggests that even with the increased number of potential targets, only distracters that belong to the same language as the targets (i.e. language match) produce a considerable congruency effect. In other words, language match between the distracter and target rather than response set membership seems to influence target identification in a word–word variant of Stroop task. Oddly, there were more errors on congruent relative to incongruent trials, which might suggest a speed-accuracy trade-off.

General discussion

In the present report, colour word target identification in a bilingual word–word Stroop task was investigated by manipulating the target word language, language match, and congruency between the distracter and target. The critical manipulation across two experiments concerned the target language. In Experiment 1, target language was manipulated between groups, with either French (L1) or English (L2) target colour words. In Experiment 2, target words from both languages were intermixed. In the blocked design

Table 7 Correlations between performance (response times and errors) and language measures

	French target condition								English target condition							
	Different language				Same language				Different language				Same language			
	Incongruent		Congruent		Incongruent		Congruent		Incongruent		Congruent		Incongruent		Congruent	
	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR
LexTALE	− .206	.049	− .050	− .113	− .092	.017	− .286	.190	− .178	− .011	− .166	.103	− .218	− .079	− .286	− .065
English Level	− *.378*	.260	− .147	.204	− .214	.158	− .283	.279	− .280	.023	− .213	*.394*	− .309	.034	− .283	.175
Years English	.084	.014	.188	− .132	.184	− .127	.217	.100	.212	− .138	.197	.015	.136	− .148	.217	− .147
% French exposure	*.417*	− *.347*	.307	− .164	*.413*	− *.363*	*.365*	− .163	**.469**	− .227	**.448**	− .263	*.413*	− .281	*.365*	− .160
% English exposure	− .114	.160	− .042	.290	− .055	.127	− .183	.119	− .156	.206	− .103	**.445**	− .108	.170	− .183	.252
FRENCH																
Acquisition	.295	− .073	*.343*	− .213	.312	− .039	.305	.037	*.359*	− .038	.300	.008	.260	− .090	.305	− .059
Fluent	.333	− .109	*.423*	− .137	**.456**	− .188	*.412*	− .068	*.378*	− .131	*.412*	.031	*.377*	− .271	*.412*	.035
Reading	***.540***	− .261	***.541***	− .124	***.568***	− .225	**.516**	.120	**.509**	− .189	**.459**	− .020	**.490**	− .211	**.516**	− .237
Fluent Reading	**.499**	− .170	**.493**	− .173	***.544***	− .232	**.449**	.148	− **.498**	− .039	**.459**	− .026	**.455**	− .160	**.449**	− .098
ENGLISH																
Acquisition	.260	− .001	.179	.131	.119	.058	.194	− .064	.168	.181	.115	− .114	.166	.035	.194	.155
Fluent	.005	− .146	.011	− .187	− .035	.018	− .028	− .168	− .047	.004	− .064	− .122	− .020	− .055	− .028	.027
Reading	.268	.013	.233	.091	.148	.177	.186	.055	.222	.212	.207	− .127	.204	.121	.186	.283
Fluent Reading	.282	− .194	.152	− .117	.154	− .077	.155	− .228	.192	.010	.110	− .060	.189	− .134	.155	− .148

Italic = $p < .05$, Bold = $p < .01$, Italic and Bold = $p < .001$. No tests were significant after Holm–Bonferroni correction

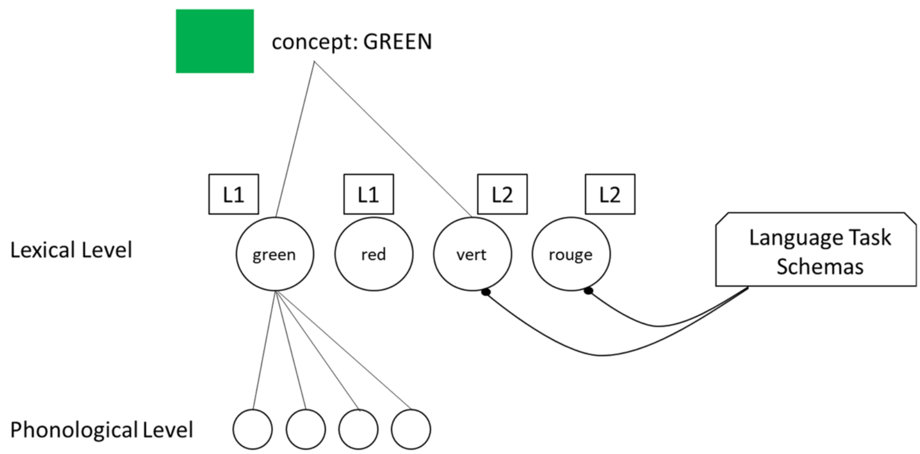

Fig. 7 A simplified version of the Inhibitory Control Model (Green, 1998). *Note*. In this example, the target language is English (L1), and a non-target language is French (L2). The inhibitory connections between the language task schemas and L2 lexical nodes indicate their suppression when L2 is not a target language

(Experiment 1), target language did not seem to matter, while when L1 and L2 occurred interchangeably (Experiment 2) as targets, responses were faster on L1.

According to the *Inhibitory Control Model* (Green 1986, 1998), illustrated at Fig. 7, each lexical representation (e.g. "green", "vert", "red", "rouge", etc.) is associated with a corresponding language tag (i.e. "English", "French"). These lexical nodes can be suppressed if they are associated with the non-target language. The semantic system (e.g. green colour concept) activates lexical nodes in both languages, but the ones from a non-target language are then suppressed reactively. This inhibition is proportional to the level of activation of the lexical nodes in the non-target language. That is, the more the semantic system activates representations in the "wrong" language, the stronger this language will be inhibited. According to the Green model, the semantic system activates L1 more strongly than L2, with a suppression being proportional to activation level. Thus, L1 should be more strongly inhibited when it is not the target language. The L2 receives less activation, and it is therefore less strongly inhibited when it is not a target language. For instance, we observed that there is nothing special about the target language in a blocked design (Experiment 1), when targets were either French (L1) or English (L2) colour words. There was no significant difference in the response speed between French and English targets, with the latter being numerically faster. This corresponds to Green's prediction of strong L1 inhibition when L1 is not a target language. For instance, in the English target group, French words occurred only as distracters, which were strongly inhibited and minimally impaired target identification. However, in Experiment 2 targets from both languages were intermixed, with L1 and L2 target colour words occurring interchangeably. With this manipulation, participants responded faster to L1 as compared to L2 words. This could be due to the fact that L1 words initially receive more activation from the semantic system that remains persistent even when L1 has to be inhibited on certain trials. Another important caveat concerns the employed response modality since the Green (1998) model is based on verbal (vocal) responses. As already mentioned, the present series of experiments made use of manual responses exclusively, which are not inherently compatible with either language. Future research might shed light on the role of response modality in Stroop word–word target identification.

As already noted, the present series of experiments made use of manual responses. Thus, the observed asymmetry might be more present if a vocal (i.e. verbal) response modality had been administered (Augustinova et al. 2019; Redding and Gerjets 1977; Sharma and McKenna 1998; White 1969). This response modality effect could be explained by different mechanisms which underlie manual and vocal responses. With manual responses, participants indicate a target word by pressing a corresponding key, while with verbal responses, participants need to name a target word aloud. In the context of the word–word Stroop task, participants would have to ignore a distracter and read the target word aloud. This involves target word recognition, but also verbal response processes, influenced by other factors, such as L2 proficiency, age of L2 acquisition, semantic context, or word frequency (Gollan et al. 2011; Thornburgh and Ryalls 1998). Different underlying processes employed during manual and verbal Stroop tasks (i.e. colour identification vs. colour naming, respectively) could account for the magnitude of congruency effect produced by each type of task (Kinoshita et al. 2017). With a manual response modality, an incongruent distracter provides evidence toward another keypress alternative. However, when vocal responses are required, the irrelevant distracter tends to activate another speech production alternative. This difference in the magnitude of the congruency effect across the two response modalities suggest that suppressing the irrelevant speech code (in a verbal Stroop) is harder than suppressing the irrelevant key response option (in a manual Stroop). That is, distracters have a strong overlearned reading association with the corresponding oral response, thus making them harder to ignore. The verbal responses (or the response modality effect in general) could be possibly integrated in the present design, when participants need to read aloud the target word, while ignoring a distracter.

Previous findings observed, however, that two words (e.g. a distracter and a target) representing members of the same semantic category (e.g. animals; "dog-pig") do little to facilitate naming of a target stimulus (Lupker 1984). A similar task was employed by Glaser and Glaser (1982; Experiment 3), who used a limited set of colour words ("red", "blue", "yellow", and "green") both as distracters and targets. When participants were instructed to name aloud the target word while ignoring the distracter, the naming latencies in the incongruent condition were longer relative to congruent and neutral conditions, with no difference between latter two (Glaser and Glaser 1982). Schmidt and colleagues (Experiment 2; 2013) used a larger set of distracter-target pairs, with target colour words preceded by either incongruent colour associates or neutral words. Participants were faster to identify a target colour word (i.e. read aloud) when it was preceded by an incongruent colour associate (e.g. "banana-green") than those preceded by neutral words (e.g. "knot-pink"; Experiment 2). It seems, therefore, that in certain conditions, incongruent primes can facilitate identification of target colour words. However, with a smaller set of repeatedly-presented stimuli, similar as in a typical Stroop task, incongruent colour word distracters interfered with identification of the target colour word (Schmidt et al. 2013).

From a bilingual perspective, the present manuscript aimed to investigate whether this congruency effect could be due to the language match between the distracter and target (same or different-language words) or due to the response set membership (whether a distracter is a potential target). To further explore the origin of this asymmetry, in Experiment 1, participants were presented with target words from only one language (either French or English). This excluded different-language distracters from being a potential target. For instance, in an English target condition, a distracter "vert" could not be a target. That is, a distracter "vert" (and other French words) is not in the response set and is therefore expected to produce a smaller congruency effect than English distracters (e.g. "green") that belong to the response set. The reverse was expected in the French target condition (Klein 1964; Risko et al. 2006; Sharma and McKenna 1998).

Our results revealed that the distracters that are potential targets (and are from the same language as target) produce larger congruency effects (e.g. "brown"–"green" is responded to slower than "green"–"green") than those which are not potential targets (and are from a different language than the target, e.g. "marron"–"green" is responded to slower than "vert"–"green"). That is, the faster responses on congruent trials suggest that to-be-ignored distracters from another language (e.g. "vert") stay salient and activate their translation equivalent (e.g. "green"), facilitating its identification. Because translation equivalents (e.g. "vert" and "green") share a common semantic representation (de Groot 1992), they are even more closely related than semantically related words within a single language (e.g. "green" and "red" or "vert" and "rouge"; Costa et al. 1999; Costa and Caramazza 1999). This is in line with the de Groot (1992) model (see *Introduction* for more details) that explains this cross-language priming by the number of semantic features shared by translation equivalents (e.g. "vert" and "green").

Experiment 2 aimed to clarify the role of the response set membership in the observed L1 and L2 asymmetry. As already discussed, all distracters, regardless of their language match with a target, served as potential targets. That is, even different-language distracters were considered as potential targets. According to this perspective, the within-subject manipulation was expected to produce larger congruency effect as compared to Experiment 1. However, almost equal response latencies between congruent and incongruent trials were observed when the distracter and target belonged to different languages (e.g. "marron"–"green" vs. "vert"–"green"), with a minimal congruency effect produced. In the same-language condition, congruent trials (e.g. "green–green" or "vert–vert") were responded to faster than incongruent trials (e.g. "brown–green" or "marron–vert", respectively). Experiment 2 therefore confirmed the notion of a larger within- than between-language congruency effect. These findings seem to align more with the language match perspective, since Experiment 2 obtained a similar pattern of results as in Experiment 1. The difference in the magnitude of between-language congruency and within-language congruency effects can be attributed to the language match between the distracter and target, rather than to the response set membership. The increased response speed on congruent trials could be due to identity priming, repeatedly reported in lexical decision literature. That is, the target classification is faster when the target is preceded by a physically identical distracter (e.g. "green–green") than by a different one (e.g. "brown–green"; Jacobs et al. 1995; La Heij et al. 1985; Perea et al. 2014; Warren 1977). Alternatively, it could be the case that this visual similarity at least partially explains speeded responses on congruent same-language trials. Further investigations are needed to clarify this issue.

Interestingly, both Experiments 1 and 2 demonstrated that the congruency effect in response latencies is modified by the language match between the distracter and target. That is, the congruency effect (i.e. the difference in response latencies between incongruent and congruent trials) is more pronounced when the distracter and target belong to the same language relative to when they belong to different languages. This confirms the notion that the within-language congruency effect is typically larger than between-language congruency effect (Kiyak 1982; MacLeod 1991; Preston and Lambert 1969).

Apart from the congruency effect, the cross-language effects could be discussed in terms of its direction. For instance, a priming effect occurs across languages in both the L2–L1 and L1–L2 directions in the lexical decision task, with the latter being reported as larger (Keatley et al. 1994; Schoonbaert et al. 2009). This larger priming in the L1–L2 direction was explained by different models of bilingual memory representation, which assume richer L1 representations (Keatley et al. 1994), stronger links to a shared conceptual store (de Groot 1992; Keatley et al. 1994; Kroll and Stewart 1994), or larger numbers of semantic nodes activated by L1 words (de Groot 1992). It is possible, however, that this priming asymmetry could be observed in certain contexts only (e.g. lexical decision task, semantic and translation priming). For instance, in the present series of experiments in which target identification was required there was no difference in L1 and L2 target identification latencies when preceded by same-language or different-language distracters. These different results reported in the lexical decision literature and the present word–word Stroop colour identification task could be due to the different contexts in which semantically related words could influence performance. For instance, an incongruent colour word distracter and a colour word target promote a *word* response in lexical decision, therefore facilitating word classification. In a Stroop identification task, incongruent distracters in either language (e.g. "brown" or "marron") indicate different response option from the one indicated by the target (e.g. "green" or "vert"). This response competition impairs target identification (Schmidt et al. 2013). Stroop response decisions depend on the evidence for each of the potential responses. In other words, evidence for a correct response is divided by evidence for other potential responses. This suggests the slower selection of correct response when a larger number of response competitors is active (Melara and Algom 2003).

Previous findings clearly show that the asymmetry between L1 and L2 congruency effect depends on the response language (Dyer 1971; Preston and Lambert 1969; Tzelgov et al. 1990). Two cross-linguistic priming directions (L1–L2 and L2–L1) could be therefore tested by manipulating a response language. For instance, if the response language matches the target language (i.e. English), participants would have to read a target word aloud (e.g. "brown"). This target identification performance could potentially be influenced by the distracter language (e.g. same or different than the target) and congruency (e.g. congruent or incongruent in meaning). In contrast, if the response language and target language are different, a target word has to be translated (i.e. "marron", brown in French). According to the Kroll and Stewart (1994) model, we should expect faster responding when an L2 target has to be identified in L1, relative to vice versa. This is due to strong lexical links from L2 to L1 that facilitate backward (L2–L1), but not forward (L1–L2) translation, which is assumed to be conceptually mediated (Kroll and Stewart 1994). Future research might aim to tease these differences further apart in both priming directions.

The impact of the automatic process of reading on the more controlled process of colour naming in a standard Stroop task (e.g. "red" in green) has been investigated across languages. For instance, this congruency (incongruent–congruent) should be stronger in L1 than in L2 due to the higher automaticity of L1 (Heidlmayr et al. 2014). This is in line with the temporal delay assumption derived from the BIA+ model (Dijkstra and van Heuven 2002), which refers to the delayed access to phonological and semantic codes in L2, relative to L1. The activation of L2 is slower, therefore producing weaker congruency effect in the Stroop task. Our data did not confirm this prediction: there was no difference in the magnitude of congruency effect between L1 and L2. According to Mägiste (1984, 1985), the amount of conflict is proportional to the mastery of the languages. In other words, the comparable size of congruency effects produced by French (L1) and English (L2) words could be due to relatively high L2 proficiency in our sample (Mägiste 1984, 1985). Future research may nevertheless aim to test this notion on a less fluent L2.

The present series of experiments compared only congruent and incongruent trials, which allowed us to measure the congruency effect exclusively. This difference in response latencies between incongruent (e.g. "green–brown") and congruent (e.g. "brown–brown") trials can be explained in terms of two possible accounts. First, according to semantic conflict account, activation of the distracter (e.g. "green") leads to inhibition of other colour concepts (e.g. target; "brown"), since both words show semantic similarity (i.e. both are colours). This semantic competition slows down target identification. Second, according to response conflict account, on incongruent trials, distracter and target activate two possible response alternatives. This conflict in the response selection stage is responsible for a delay in responding. Both types of conflict occur for L2 words (Šaban and Schmidt 2021; Schmidt et al. 2018). Future research might aim to dissociate stimulus and response conflict in both language match and language mismatch conditions.

Future research could also integrate a neutral condition (e.g. letter strings such as "xxxx", or colour-neutral words in L1 and L2), in which target colour word is preceded by colour-neutral distracters. Faster responses in the congruent relative to the control trials indicates a facilitation effect. Slower responses in the incongruent relative to the neutral trials indicate an interference effect. Facilitation effects are typically much smaller than the interference effect (MacLeod 1998). Future work might also explore facilitation and interference effects in both language match and language mismatch conditions. As another interesting aside, the "word–word" Stroop task variant is not limited to the use of colour-related stimuli as in a standard colour-word Stroop procedure. It can be used with any word type, therefore allowing the exploration of a larger scope of cross-linguistic semantic and associative relationships. As such, the "word–word" variant of the Stroop task is more similar to the priming tasks that are typically used in a large number of semantic domains (Fischler 1977; Glaser and Glaser 1982; Neely 1977; Schmidt et al. 2013).

Conclusion

The present series of experiments suggests that there is a certain overlap in semantic activation produced by L1 and L2 words. That is, instead of depending heavily on the target language or response-set membership, the congruency effect mostly depends on the language match between the distracter and target in our word–word Stroop task. Only under certain conditions, a target identification is favourited in L1 relative to L2. The present work is a good starting point in exploring the word–word Stroop target identification task on different word types. Moreover, it is recognized as suitable for conducting further investigations of bilingual semantic activation.

Declarations

Conflict of interests The authors have no conflicts of interest to declare.

Ethical approval In France, the "Loi Jardé" specifies that Ethic review and approval are necessary for research that involved human participants only if the research aims at developing biological or medical knowledge (which is not the case of our study), and that the experimentations in Human Sciences (even in the health domain) are not falling within this scope. Therefore, our study did not require an ethic approval. However, even if our study does not require an ethic approval according to the French law, it was conducted in accordance with the Declaration of Helsinki (1964) and each participant provided written informed consent. The link of the "Loi Jardé", and the specific part of the "Loi Jardé" that gives the scope of the ethical review and approval conditions: https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000034634225

References

Altarriba J, Mathis KM (1997) Conceptual and lexical development in second language acquisition. J Mem Lang 36(4):550–568. https://doi.org/10.1006/jmla.1997.2493

Augustinova M, Parris BA, Ferrand L (2019) The loci of Stroop interference and facilitation effects with manual and vocal responses. Front Psychol 10:1786. https://doi.org/10.3389/fpsyg.2019.01786

Bialystok E, Craik F, Luk G (2008) Cognitive control and lexical access in younger and older bilinguals. J Exp Psychol Learn Mem Cogn 34(4):859–873. https://doi.org/10.1037/0278-7393.34.4.859

Brauer M (1998) Stroop interference in bilinguals: the role of script similarity between two languages. In: Healy AF, Bourne LE Jr (eds) Foreign language learning: psycholinguistic studies on training and retention. Erlbaum, pp 317–337

Chen H-C, Leung Y-S (1989) Patterns of lexical processing in a non-native language. J Exp Psychol Learn Mem Cogn 15(2):316–325

Costa A, Caramazza A (1999) Is lexical selection in bilingual speech production language-specific? Further evidence from Spanish–English and English–Spanish bilinguals. Biling Lang Cogn 2(3):231–244. https://doi.org/10.1017/S1366728999000334

Costa A, Miozzo M, Caramazza A (1999) Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? J Mem Lang 41(3):365–397. https://doi.org/10.1006/jmla.1999.2651

Dalrymple-Alford EC, Budayr B (1966) Examination of some aspects of the Stroop color-word test. Percept Mot Skills 23(3):1211–1214. https://doi.org/10.2466/pms.1966.23.3f.1211

de Groot AM (1992) Determinants of word translation. J Exp Psychol Learn Mem Cogn 18(5):1001–1018. https://doi.org/10.1037/0278-7393.18.5.1001

Dijkstra T, van Heuven WJB (2002) The architecture of the bilingual word recognition system: from identification to decision. Biling Lang Cogn 5(3):175–197. https://doi.org/10.1017/S1366728902003012

Dyer FN (1971) Color-naming interference in monolinguals and bilinguals. J Verbal Learn Verbal Behav 10(3):297–302. https://doi.org/10.1016/S0022-5371(71)80057-9

Fang S-P, Tzeng OJL, Alva L (1981) Intralanguage vs. interlanguage Stroop effects in two types of writing systems. Mem Cogn 9(6):609–617. https://doi.org/10.3758/BF03202355

Fischler I (1977) Semantic facilitation without association in a lexical decision task. Mem Cogn 5(3):335–339. https://doi.org/10.3758/BF03197580

Glaser MO, Glaser WR (1982) Time course analysis of the Stroop phenomenon. J Exp Psychol Hum Percept Perform 8(6):875–894

Glaser MO, Glaser WR (1989) Context effects in Stroop-like word and picture processing. J Exp Psychol Gen 118:13–42

Gollan TH, Montoya RI, Cera C, Sandoval TC (2009) More use almost always a means a smaller frequency effect: aging, bilingualism, and the weaker links hypothesis. J Mem Lang 58(3):787–814

Gollan TH, Slattery TJ, Goldenberg D, Van Assche E, Duyck W, Rayner K (2011) Frequency drives lexical access in reading but not in speaking: the frequency-lag hypothesis. J Exp Psychol Gen 140(2):186–209. https://doi.org/10.1037/a0022256

Green DW (1986) Control, activation, and resource: a framework and a model for the control of speech in bilinguals. Brain Lang 27:210–223

Green DW (1998) Mental control of the bilingual lexico-semantic system. Biling Lang Cogn 1(2):67–81. https://doi.org/10.1017/S1366728998000133

Heidlmayr K, Moutier S, Hemforth B, Courtin C, Tanzmeister R, Isel F (2014) Successive bilingualism and executive functions: the effect of second language use on inhibitory control in a behavioural Stroop Colour Word task. Biling Lang Cogn 17(3):630–645. https://doi.org/10.1017/S1366728913000539

Jacobs AM, Grainger J, Ferrand L (1995) The incremental priming technique: a method for determining within-condition priming effects. Percept Psychophys 57(8):1101–1110. https://doi.org/10.3758/BF03208367

Keatley CW, Spinks JA, De Gelder B (1994) Asymmetrical cross-language priming effects. Mem Cogn 22(1):70–84. https://doi.org/10.3758/BF03202763

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. J Exp Psychol Learn Mem Cogn 43(3):369–384. https://doi.org/10.1037/xlm0000311
Kiyak HA (1982) Interlingual interference in naming color words. J Cross Cult Psychol 13(1):125–135
Klein GS (1964) Semantic power measured through the interference of words with color-naming. Am J Psychol 77(4):576. https://doi.org/10.2307/1420768
Kroll JF, Stewart E (1994) Category interference in translation and picture naming: evidence for asymmetric connections between bilingual memory representations. J Mem Lang 33:149–174
La Heij W, Van der Heijden AHC, Schreuder R (1985) Semantic priming and Stroop-like interference in word-naming tasks. J Exp Psychol Hum Percept Perform 11(1):62–80
Lemhöfer K, Broersma M (2012) Introducing LexTALE: a quick and valid lexical test for advanced learners of English. Behav Res Methods 44(2):325–343. https://doi.org/10.3758/s13428-011-0146-0
Logan GD, Zbrodoff NJ (1979) When it helps to be misled: facilitative effects of increasing the frequency of conflicting stimuli in a Stroop-like task. Mem Cogn 7(3):166–174. https://doi.org/10.3758/BF03197535
Lupker SJ (1984) Semantic priming without association: a second look. J Verbal Learn Verbal Behav 23(6):709–733
MacLeod CM (1991) Half a century of research on the Stroop effect: an integrative review. Psychol Bull 109(2):163–203. https://doi.org/10.1037/0033-2909.109.2.163
MacLeod CM (1998) Training on integrated versus separated Stroop tasks: the progression of interference and facilitation. Mem Cogn 26(2):201–211. https://doi.org/10.3758/BF03201133
Mägiste E (1984) Learning a third language. J Multiling Multicult Dev 5(5):415–421. https://doi.org/10.1080/01434632.1984.9994170
Mägiste E (1985) Development of intra- and interlingual interference in bilinguals. J Psycholinguist Res 14(2):137–154. https://doi.org/10.1007/BF01067626
Marian V, Blumenfeld HK, Kaushanskaya M (2007) The Language Experience and Proficiency Questionnaire (LEAP-Q): assessing language profiles in bilinguals and multilinguals. J Speech Lang Hear Res 50(4):940–967. https://doi.org/10.1044/1092-4388(2007/067)
Melara RD, Algom D (2003) Driven by information: a tectonic theory of Stroop effects. Psychol Rev 110(3):422–471. https://doi.org/10.1037/0033-295X.110.3.422
Neely JH (1977) Semantic priming and retrieval from lexical memory: roles of inhibitionless spreading activation and limited-capacity attention. J Exp Psychol Gen 106(3):226–254
Paivio A, Clark JM, Lambert WE (1988) Bilingual dual-coding theory and semantic repetition effects on recall. J Exp Psychol Learn Mem Cogn 14(1):163–172
Perea M, Jiménez M, Gómez P (2014) A challenging dissociation in masked identity priming with the lexical decision task. Acta Physiol (oxf) 148:130–135. https://doi.org/10.1016/j.actpsy.2014.01.014
Potter MC, So K-F, Eckardt BV, Feldman LB (1984) Lexical and conceptual representation in beginning and proficient bilinguals. J Verbal Learn Verbal Behav 23(1):23–38. https://doi.org/10.1016/S0022-5371(84)90489-4
Preston MS, Lambert WE (1969) Interlingual interference in a bilingual version of the Stroop color-word task. J Verbal Learn Verbal Behav 8(2):295–301. https://doi.org/10.1016/S0022-5371(69)80079-4
Redding GM, Gerjets DA (1977) Stroop effect: interference and facilitation with verbal and manual responses. Percept Mot Skills 45(1):11–17. https://doi.org/10.2466/pms.1977.45.1.11
Risko EF, Schmidt JR, Besner D (2006) Filling a gap in the semantic gradient: color associates and response set effects in the Stroop task. Psychon Bull Rev 13(2):310–315. https://doi.org/10.3758/BF03193849
Šaban I, Schmidt JR (2021) Stimulus and response conflict from a second language: Stroop interference in weakly-bilingual and recently-trained languages. Acta Physiol (oxf) 218:103360. https://doi.org/10.1016/j.actpsy.2021.103360
Schmidt JR, Besner D (2008) The Stroop effect: Why proportion congruent has nothing to do with congruency and everything to do with contingency. J Exp Psychol Learn Mem Cogn 34(3):514–523. https://doi.org/10.1037/0278-7393.34.3.514
Schmidt JR, Cheesman J, Besner D (2013) You can't Stroop a lexical decision: Is semantic processing fundamentally facilitative? Can J Exp Psychol/revue Canadienne De Psychologie Expérimentale 67(2):130–139. https://doi.org/10.1037/a0030355
Schmidt JR, Hartsuiker RJ, De Houwer J (2018) Interference in Dutch-French bilinguals: stimulus and response conflict in intra- and interlingual Stroop. Exp Psychol 65(1):13–22. https://doi.org/10.1027/1618-3169/a000384
Schoonbaert S, Duyck W, Brysbaert M, Hartsuiker RJ (2009) Semantic and translation priming from a first language to a second and back: making sense of the findings. Mem Cogn 37(5):569–586. https://doi.org/10.3758/MC.37.5.569
Sharma D, McKenna FP (1998) Differential components of the manual and vocal Stroop tasks. Mem Cogn 26(5):1033–1040. https://doi.org/10.3758/BF03201181
Stoet G (2010) PsyToolkit: a software package for programming psychological experiments using Linux. Behav Res 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. Teach Psychol 44(1):24–31. https://doi.org/10.1177/0098628316677643
Stroop JR (1935) Studies on interference in serial verbal reactions. J Exp Psychol 18:643–661
Thornburgh DF, Ryalls JH (1998) Voice onset time in Spanish–English bilinguals: early versus late learners of English. J Commun Disord 31(3):215–229. https://doi.org/10.1016/S0021-9924(97)00053-1
Tzelgov J, Henik A, Leiser D (1990) Controlling Stroop interference: evidence from a bilingual task. J Exp Psychol Learn Mem Cogn 16(5):760–771. https://doi.org/10.1037/0278-7393.16.5.760
Warren RE (1977) Time and the spread of activation in memory. J Exp Psychol Hum Learn Mem 3(4):458–466
White BW (1969) Interference in identifying attributes and attribute names. Percept Psychophys 6(3):166–168. https://doi.org/10.3758/BF03210086

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.