Audio-visual Stroop Matching Task with First and Second Language Colour Words and Colour Associates

Iva Šaban and James R. Schmidt

LEAD-CNRS UMR 5022, Université de Bourgogne, Dijon

Author notes

This work was supported by the French "Investissements d'Avenir" program, project ISITE-BFC (contract ANR15-IDEX-0003) to James R. Schmidt. Replication package is available on the Open Science Framework (https://osf.io/48q2p/?view_only=3c4cdec3f832446984291fc5f22f6392).

Abstract

In the audio-visual Stroop matching task, participants compare one Stroop stimulus dimension (e.g., the colour of a written word) to a second stimulus (e.g., a spoken word) and indicate whether these two stimuli match or mismatch. Slower responses on certain trials can be due to conflict which occurs between colour representations (*semantic conflict*) or due to conflict between responses evoked by task comparisons (*response conflict*). The contribution of these conflicts has been investigated with colour word distracters. This is the first study which explores how two types of first and second language words affect audio-visual matching. Native French speakers performed a bilingual Stroop matching task with intermixed French (L1) and English (L2) colour words (Experiment 1) and colour associates (Experiment 2) presented in congruent and incongruent colours simultaneously with spoken French colour words. Participants were instructed to indicate whether the spoken word “matches” or “mismatches” the font colour, while ignoring written word meaning. Interestingly, the results were similar for the critical “mismatch” trials for both French and English words. The responses were the fastest on trials in which task comparisons activate fewer response alternatives, supporting the assumption of the response conflict account.

Keywords: audio-visual matching, between-language interference, within-language interference, semantic conflict, response conflict

Cognitive control measured by the Stroop task and corresponding conflict effects

People make everyday decisions about allocating cognitive control in order to pursue their goals (e.g., what to pay attention to, what to stop themselves from doing). For instance, when confronted with multiple sources of information, our cognitive system adapts our attentional resources away from distracting (i.e., non-goal relevant) stimuli and/or toward the goal-relevant stimuli and the action we are supposed to make. The Stroop task is one particularly useful tool in assessing the ability of the cognitive control system to control selective attention. In the Stroop task, participants are instructed to name the ink colour of the written word while ignoring its meaning. The standard finding of slower and less accurate responding on incongruent (e.g., “red” in green) relative to congruent (e.g., “red” in red) trials is known as the *congruency* or *Stroop effect* (Stroop, 1935; for a review, see MacLeod, 1991). Among other things, the Stroop effect indicates that control over selective attention is not absolute: the distracting word influences colour naming, indicating that it is not ignored entirely.

One other question of interest concerns the source of this congruency effect. According to *response conflict accounts,* word reading and colour naming compete for a single response channel (Goldfarb & Henik, 2007; Morton, 1969; Posner & Snyder, 1975). The word reading response becomes available prior to a colour naming response, because it is a faster and more automatized process than colour naming (for the automaticity of reading debate, see Augustinova & Ferrand, 2014; Besner et al., 1997). Thus, word reading disrupts colour naming but not vice versa. Alternatively, *semantic (or stimulus) conflict accounts* assume that the conflict occurs in an earlier phase of processing (Luo, 1999; Seymour, 1977; Simon & Berbaum, 1988). When the ink colour and word meaning are incongruent (e.g., “red” in green), two distinct semantic representations (“red” and “green”) are simultaneously activated. This semantic conflict takes time to resolve, presumably before response selection.

Various authors have discussed the relative contribution of semantic and response conflict in explaining the source of congruency. Nowadays, the current consensus is that both effects contribute to the standard Stroop effect (Ferrand & Augustinova, 2014). The presence of semantic and response conflict indicates that the distracting word slipped through the attentional filter, either at an early semantic processing phase, or later response selection phase. Most models (Glaser & Glaser, 1989) assume that semantic processing occurs earlier in the stimulus processing, with the response being selected at a later stage.

Stroop matching task

In a Stroop task, a to-be-ignored written word stimulus and the oral response (e.g., colour naming and word reading) are compatible, which has been suggested as an inherent limitation of the Stroop task (Treisman & Fearnley, 1969). That is, a response in the form of a spoken word is required in both colour naming and word reading tasks. This might produce a congruency effect only when the irrelevant stimulus attribute (e.g., word) belongs to the same class as the response. This limitation has inspired a novel variant of the Stroop task, named the *Stroop matching task*, in which responses are neither words nor colours.

In the Stroop matching task, participants are instructed to make matching/mismatching judgements on two simultaneously presented stimuli (Treisman & Fearnley, 1969). That is, participants are asked to indicate whether two stimulus dimensions “match” or “mismatch” (e.g., two colour words or a word and colour). Most importantly, this task permits a test of the contribution of two contrasting potential sources of conflict: *semantic* and *response conflict*. For instance, in the *meaning decision task* of Dyer (1973), participants were asked to compare a colour word to a colour patch and to ignore the print colour of the word.

Matching/mismatching judgements were slower when the colour word was printed in an incongruent colour. However, responses are slower to “match” trials when the word mismatches the colour (e.g., “red” in blue) than when the word and colour match (e.g., “red” in red). This is because the incongruent colour activates a semantic representation (i.e., blue) that competes with the representations activated by the other stimuli (i.e., red). According to this perspective, then, semantic conflict interferes with the matching/mismatching response (Dyer, 1973; Flowers, 1975). This finding challenges the assumptions of certain response conflict accounts because the supposedly slower colour naming response (i.e., “blue”) influenced responding more than the faster word meaning response (i.e., “red”).

Similar findings were observed with the *visual decision task* in which participants are asked to decide whether two stimuli have the same ink colour (Egeth et al., 1969; Virzi & Egeth, 1985). For instance, on a trial with the word “red” printed in blue and a blue patch, the required response is “match”. Interestingly, the conflicting verbal information provided by the word (i.e., “red”) did not produce interference, seemingly indicating that the word meaning is not fast enough to compete with the semantic unit (“blue”) accessed by the word’s ink colour (Egeth et al., 1969; Treisman & Fearnley, 1969). This finding again contradicts the assumptions of the response conflict account, since word reading, although faster than colour naming, produced no interference with responding. However, when the colour names were replaced with the words “SAME” and “DIFF”, interference reappeared. That is, two simultaneously presented words “DIFF” printed in the same colour (e.g., red) resulted in interference, because the correct response for the colours (i.e., “matching” or “SAME”) competes with the response suggested by the distracters (i.e., “mismatching” or “DIFF”). This indicates that participants had difficulties to ignore the written words and respond to the ink colour exclusively, as assumed by the response conflict account (Egeth et al., 1969).

The *meaning decision* and *visual decision* tasks have been integrated within a single matching procedure to directly test whether interference is due to semantic or response conflict. Luo (1999) replicated both the interference in the meaning decision task and the absence of interference in the visual decision task. Luo argued that only the meaning decision task required participants to access the semantic system. In this task, when a Stroop stimulus “red” printed in blue is presented with a red patch (i.e., “matching” response is required), the ink colour and the colour patch activate two competing semantic representations (e.g., “blue” and “red”). According to Luo (1999), this generates a semantic conflict. In contrast, these findings are difficult to explain by the response conflict account because it did not matter whether the response was “matching” or “mismatching” since the response latencies were faster for related ink colours than for unrelated ink colours.

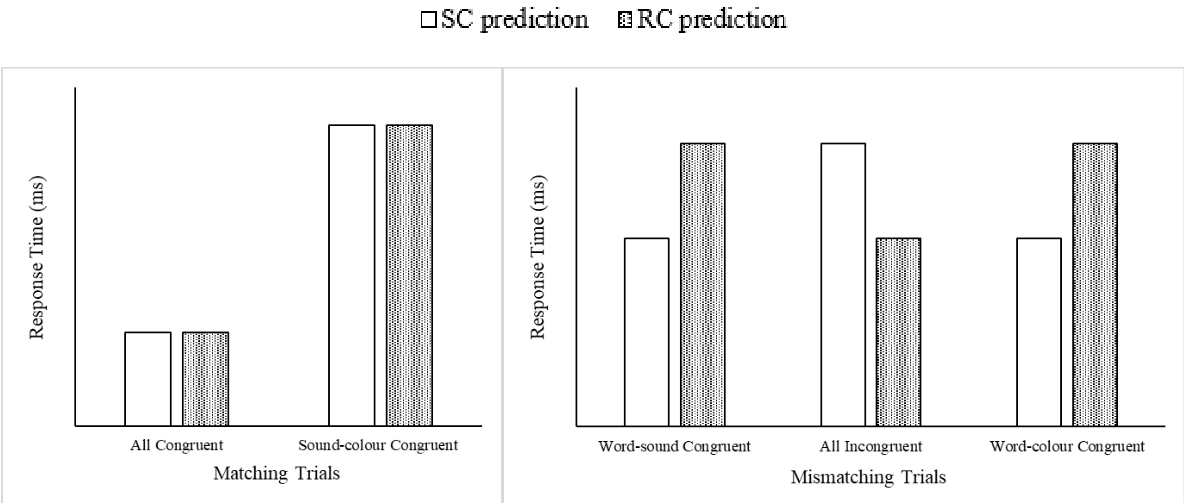
However, Goldfarb and Henik (2006) pointed out that Luo’s (1999) analysis on the *meaning decision task* only distinguished between a “mismatching” condition in which coloured patches appeared together with either an incongruent colour word (e.g., “red” in blue paired with a blue rectangle) or a congruent colour word (e.g., “red” in red paired with a blue rectangle). Goldfarb and Henik suggested that the congruency of the colour word stimuli could play a role in producing a conflict. For both “matching” and “mismatching” responses, Stroop stimuli could be either congruent or incongruent. Thus, in addition to the four conditions contrasted by Luo (1999), Goldfarb and Henik (2006) introduced a condition in which both dimensions of the incongruent Stroop stimuli mismatch with the colour of the patch (e.g., “red” in blue with a green patch). They observed that “matching” responses were faster when Stroop stimuli were congruent (e.g., “red” in red with a red patch) than when they were incongruent (e.g., “red” in green with a red patch). The “mismatching” responses were the slowest when the word and ink colour were congruent (e.g., “red” in red with a green patch). Delays were similar when the ink colour and patch colour matched (e.g., “red” in green with a green patch) and when they mismatched (e.g., “red” in blue with a green patch). To sum up, response latencies to incongruent trials were slower during “matching” responses and faster during “mismatching” responses. According to Goldfarb and Henik, participants erroneously made an irrelevant match between the word and its ink colour. That is, seeing congruent and incongruent Stroop stimuli leads to a covert "matching" and "mismatching" response, respectively, which can either facilitate or interfere with the actual response required. Thus, they suggested that the results are clearly in line with the response conflict account.

In a related matching task variant, Bornstein (2015) asked participants to make an audio-visual matching judgement based on the task-relevant auditory (i.e., spoken colour word) and visual stimuli (i.e., ink colour of a written word). On each trial, participants were instructed to indicate whether the colour of a written word (while ignoring its meaning) corresponds to a simultaneously presented spoken word. Bornstein (2015) compared the interference produced by congruent and incongruent written stimuli on matching spoken word and font colour. Bornstein observed that incongruent distracters (e.g., "red" in blue while hearing "blue") interfered more than congruent distracters (e.g., "blue" in blue while hearing "blue") with "matching" responses, similarly to Goldfarb and Henik (2006). Furthermore, written words that were congruent with either task-relevant dimension (i.e., ink colour or spoken word) interfered with "mismatching" responses relative to trials in which the word mismatched both (e.g., "green" in red while hearing "blue").

Both the semantic and response conflict accounts assume the same outcome for "matching" responses with faster responses on congruent (i.e., *All congruent*) relative to incongruent colour words (i.e., *Sound-colour congruent*). According to the semantic conflict account, this is due to the fact that for congruent colour words, all three task dimensions refer to the same colour (i.e., blue). The response conflict account explains this difference in response speed by the three stimulus comparisons, which all suggest the same response alternative (i.e., "match"). Critically, the assumptions of these two accounts differ for "mismatching" trials. According to the semantic conflict account, *All incongruent* trials, in which a written colour word is incongruent (e.g., “green” in red, hear “blue”) with the remaining two colour dimensions, should produce the largest interference. Three different semantic representations (i.e., blue, red, and green) are simultaneously activated, thus slowing down responding. In contrast, the response conflict account suggests that incongruent colour word distracters should facilitate responding when both dimensions (e.g., green and red) are incompatible with a spoken word (e.g., blue). This is because all three comparisons (i.e., written vs. spoken word, written word vs. colour, and spoken word vs. colour) provide evidence toward the same response alternative (i.e., “mismatching”), resulting in faster response latencies (Bornstein, 2015; Caldas et al., 2012; Goldfarb & Henik, 2006). The shared prediction of semantic and response conflict accounts for “matching” trials and contrasting predictions for “mismatching” trials are visualised in Figure 1.

Figure 1

Prediction of semantic and response conflict accounts for “matching” and “mismatching” trials



Colour Associates

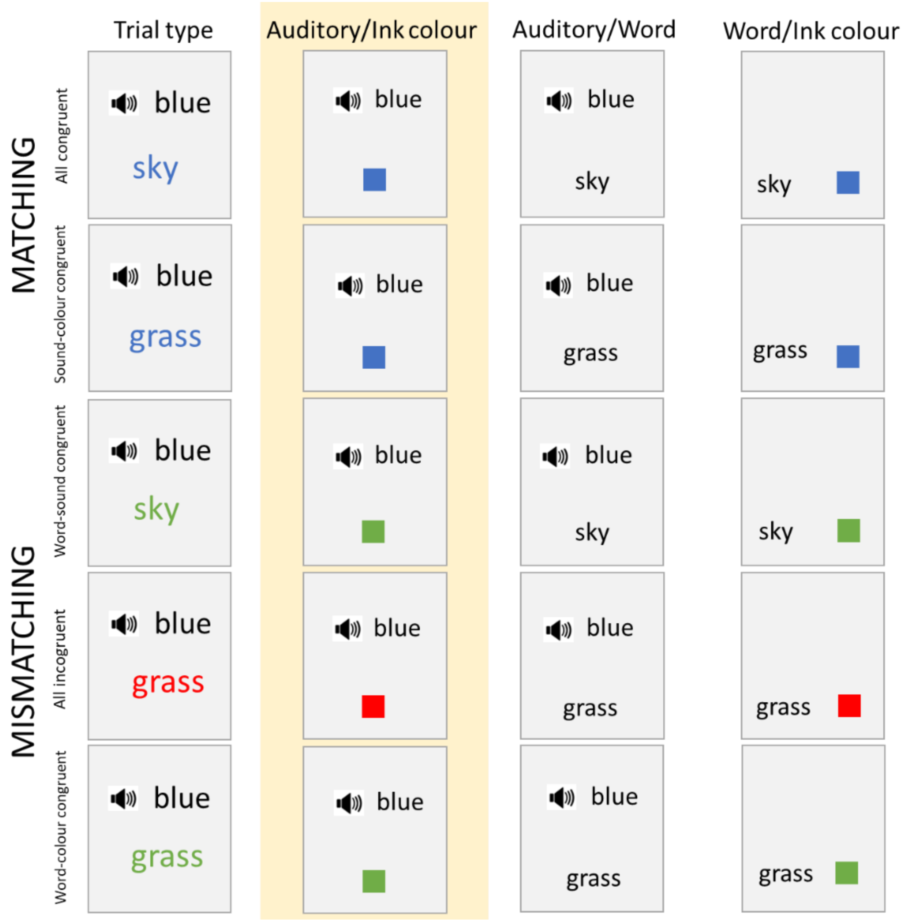
All previously described Stroop matching task studies made use of colour words. However, similar studies have not been conducted with another common word type with a strong colour dimension, namely, colour associates, which could help further evaluate conflict effects in the Stroop matching task. Colour associates are words that are closely related to colour words (e.g., “sky” with blue) and their semantic representations (Tanaka & Presnell, 1999). Colour associates do produce interference with colour naming in the Stroop task. Similar to colour words, colour associates can be congruent (e.g., “sky” in blue) or incongruent (e.g., “sky” in red) with the ink colour. When contrasting the response latencies of these two types of trials, a congruency occurs, with slower and less accurate responses on incongruent relative to congruent colour associates (Glaser & Glaser, 1989; Klein, 1964; Risko et al., 2006; Schmidt & Cheesman, 2005).

This difference in performance might be due to early semantic processes (Glaser & Glaser, 1989). When a colour word distracter is printed in an incongruent colour (e.g., “sky” in red), two competing colour representations (i.e., red and blue) are simultaneously activated, thus producing semantic conflict. According to this perspective, colour associate congruency effects arise from early, semantic processes. Another account suggests that colour associates might directly produce the colour response linked to the colour associate. That is, when the word “sky” is printed in red, both the responses linked to the colour blue (i.e., the colour associated with “sky”) and the response linked to the colour red (i.e., which is associated to the ink colour) will be activated. Thus, according to this perspective, incongruent colour associates produce response competition, resulting in response conflict exclusively, rather that semantic conflict (Klein, 1964). Third, Sharma and McKenna (1998) suggested that interference should occur only when vocal responses are required and should be eliminated with manual responses, though subsequent research clearly indicates the presence of conflict effects in keypress tasks (e.g., Schmidt & Cheesman, 2005).

One reason why colour associates might be especially interesting in the context of the matching task relates to a peculiarity of the matching task. For “matching” trials, both the semantic and response conflict accounts make identical predictions. For "mismatching" trials, the two accounts make exactly opposite predictions. Specifically, the semantic conflict account suggests that *All incongruent* trials should be slower than the two other types of "mismatching" trial types, whereas the response conflict account suggests that *All incongruent* trials should be faster than the two other types of "mismatching" trial types. Therefore, if *both* semantic and response conflict occur, the larger of the two effects will "mask" the other. In particular, evidence of a response conflict effect *could* indicate that only response conflict occurs in the matching task but could also indicate that response conflict is merely larger than semantic conflict. Thus, if the response conflict effect can be eliminated, then we might expect that the "true" effect of semantic conflict would be revealed. Although some competing accounts of colour associates conflict effects exist (as discussed above), we hypothesized that colour associates would produce only semantic conflict. Some evidence suggests this to be the case in standard Stroop studies (e.g., Schmidt & Cheesman, 2005). All task comparisons (one relevant and two irrelevant) for each colour associate trials are visualised in Figure 2.

Figure 2

Types of trials and example stimuli with relevant (highlighted column) and irrelevant task comparisons



Bilingualism

The Stroop effect has been frequently investigated in bilingual people (Altarriba & Mathis, 1997; Dyer, 1971; MacLeod, 1991; Mägiste, 1982; Preston & Lambert, 1969; Tzelgov et al., 1990). These previous studies showed that congruency can be observed with both first language (L1) and second language (L2) words. However, the interference is generally larger for L1 words than for L2 words. This could be explained by the nature of L2 connections. For instance, there has been debate about whether L2 words 1) have strong direct connections to semantic representations but weak connections to the L1 lexicon, 2) are strongly connected to the L1 lexicon but not semantics, or 3) have both semantic and lexical connections (Altarriba & Mathis, 1997; Kroll & Stewart, 1994; Schmidt et al., 2018). Thus, it is unclear whether L2 words would lead to semantic conflict, response conflict, or a combination of both. Specifically, L2 words would not be expected to generate semantic conflict if they have no (or very weak) connections to semantics. If the exact reverse is true and L2 words function as semantic associates to their L1 translations, then only semantic conflict might be expected, as discussed in the previous section on colour associates.

Another important question in the bilingual Stroop literature concerns the modulation of Stroop interference by stimulus and response language (i.e., the language of a distracter and the language of a response, respectively). First, the distracter language can match the response language. For instance, colour naming of the distracter “red” printed in green produces within-language (or intralingual) interference when English is a response language (i.e., a correct response is to say “green”). Second, the distracter language can mismatch the response language. That is, colour naming of the distracter “rouge” (red in French) printed in blue produces between-language (or interlingual) interference when English is a response language (i.e., a correct response is to say “green”).

The magnitude of within- and between-language interference has been compared repeatedly. A standard finding is a larger within-language than between-language interference effect (Dyer, 1971; Hamers & Lambert, 1972; Kiyak, 1982; MacLeod, 1991; Preston & Lambert, 1969). For instance, MacLeod (1991) reported that the between-language interference represents about 75% of within-language interference. However, these findings mostly originated from the standard visual (MacLeod, 1991) and auditory (Hamers & Lambert, 1972) Stroop task but have never been confirmed with the Stroop matching task. In a bilingual Stroop matching task, it might be assumed that distracters that match in language with a spoken word will produce larger interference relative to those that mismatch. To test this in the present series of studies, we used distracting words from both the first language (i.e., French) and a second language (i.e., English). However, spoken words were always French. French distracters are therefore expected to produce larger interference (i.e., within-language interference) relative to English distracters (e.g., between-language interference).

Present Study

In the present series of experiments a bilingual audio-visual Stroop matching task was designed to further explore the 1) magnitude of interference produced by first (L1) and second (L2) language colour words and colour associates, and 2) the relative contributions of semantic and response conflict. In addition to first language colour words, frequently used as distracters in the literature, we introduced second-language colour words (Experiment 1). That is, intermixed French (L1) and English (L2) colour words served as distracters, while participants had to match its ink colour with a spoken French colour word. Thus, this manipulation allows us to test the consensus of larger within- than between-language interference. If this is the case, a larger interference effect is expected to occur with French (L1) than with English (L2) colour word distracters. The design of this study can be found in the *Audiovisual Stimulus Combination* section. Experiment 2 aims to further expand the findings by using colour associates instead of colour words. That is, both French and English colour associates were used as distracters, with participants matching their ink colour with a spoken French colour word. Note that, in contrast to Experiment 1, a spoken word (e.g., “vert”, French for green) does not correspond to a written word (e.g., “herbe”, French for grass). This manipulation should (according to some views) eliminate response conflict since “herbe” might be unable to retrieve the response linked to green. Furthermore, this could reveal the role of the semantic conflict, which is possibly masked by a (larger) response conflict effect. Apart from that, the question of larger within- relative to between-language interference remains open. That is, French colour-associates are expected to produce more interference than their English counterparts.

The present series of studies also aims to investigate the source of this interference. As already discussed, the interference could be due to the conflict between semantic representations (i.e., semantic conflict) or due to the conflict between response alternatives (i.e., response conflict). Based on the findings of Luo (1999) and Goldfarb & Henik (2003), these two opposing accounts predict similar outcomes for “matching” responses. That is, when a correct response is “match”, *Sound-colour congruent* trials will produce slower responses than *All congruent* distracters. However, semantic- and response-conflict accounts make different assumptions for “mismatching” responses, based on the congruency between task dimensions. According to the semantic conflict account, a written distracter should produce the largest interference by being incongruent with both task dimensions (e.g., on *All incongruent* trials) than by being incongruent with only one of them (e.g., on *Word-sound congruent* and *Word-colour congruent* trials). This is because, on *All incongruent* trials, the distracting written word is incongruent with both target dimensions, thus producing a delay in responding. In contrast, the response-conflict account assumes that the smallest interference will be observed with *All incongruent* trials, when all task comparisons suggest the same, “mismatching” response. That is, interference will be mostly observed on *Word-sound congruent* and *Word-colour congruent* trials, where one of the irrelevant task comparisons suggest the same response alternative as the relevant comparison (i.e., “mismatch”), but the third comparisons suggest the other (incorrect) response alternative (i.e., “match”).

Experiment 1

Experiment 1 contrasts the response latencies on congruent and incongruent French (L1) and English (L2) colour word distracters, each accompanied by a French spoken word.

Participants were instructed to respond according to whether the ink colour and spoken word match or mismatch by pressing the corresponding key. The combinations of visual and auditory stimuli produced five trial types: two "matching" and three "mismatching", discussed in detail in the *Audiovisual Stimulus Combination* section. The aim of Experiment 1 was to 1) compare the magnitude of interference produced by first and second language colour words in the audio-visual Stroop matching task, and 2) investigate whether this interference is due to semantic or response conflict.

Method

Participants

A total of 34 (31 women) [removed for review] undergraduates (M_{age} = 19; *SD* = .78) voluntarily participated in the experiment in exchange for course credit. An a priori power analysis was conducted using G*Power 3 (Faul et al., 2007) for sample size estimation, based on data from Goldfarb and Henik (2006), N=12, which compared response times on matching and mismatching trials separately. The effect size in Goldfarb and Henik's (2006) study was $\eta_p^2 = .57$, considered to be large. With a significance criterion of $\alpha = .05$ and power .95, the minimum sample size needed with this effect size is N = 22 for repeated measures ANOVA. Preferring more power than minimally necessary, we decided to collect data for at least 30 participants, stopping after a testing week when this number was exceeded (resulting in the obtained sample size of N = 34).

All participants had normal of corrected-to-normal visual acuity, normal colour vision and normal auditory acuity, as assessed via screening questions. Participants gave written informed consent before the study. All the procedures were conducted in accordance with the Declaration of Helsinki, although nonbiomedical research in [removed for review] does not require ethics approval. All participants were native French speakers. A language questionnaire (to be discussed shortly) was used to assess and confirm that participants fit with these criteria. Average language background scores (mean age and standard errors) are presented in Table 1 (see Results section).

Apparatus

The experiment was conducted in a sound-attenuated room in the laboratory. Stimulus presentation and response timing were controlled and recorded by Psytoolkit (Stoet, 2010, 2017). The study was conducted using a PC laptop with an AZERTY keyboard and a 15’’ monitor. Participants responded with the “D” key when the audio and the ink colour of the written distracted mismatched (e.g., hear “green” and see “brown” in brown). Participants responded with the “K” key when the audio and the ink colour matched (e.g., hear “green” and see “brown” in green). Prior to the Stroop matching portion of the experiment, participants filled out a short language demographic questionnaire. This questionnaire asked for gender, age, native language, years of English training in school, a self-rating of English knowledge ranging from 0 (= almost none) to 5 (= perfect). A subset of questions from the French version of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian et al., 2007) was inserted. In particular, the questions asking participants to list the languages in order of dominance and acquisition were retained. They were also asked to indicate the percentage with which they used French and English in the recent period. Also retained from the LEAP-Q were two boxes, one for French and one for English, asking for the age the participants 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 this questionnaire was to assure that participants had the correct language dominance. Finally, as an addition to these two questionnaires, participants were asked to give the French translations of the four English words used in the experiment (i.e., “green”, “brown”, “pink” and “white”).

This was followed by the LexTale English vocabulary test (Lemhöfer & Broersma, 2012) with instructions translated into French. This test contains 63 English-looking words (3 practice trials and 60 test trials). 2/3 of the test trials are actual English words (e.g., “moonlit”, “fluid”), whereas the remaining 1/3 are not (e.g., “plaudate”, “rebondicate”). Participants were instructed to select the words that they are certain are actual English words. Correct “hits” were rewarded with one point, and incorrect “false alarms” were penalized by two points.

Materials and design

During the experimental part of the experiment, participants were presented with a set of French-English translation equivalents (i.e., “green/vert”, “brown/marron”, “rose/pink”, and “white/blanc”), typed in lowercase Courier New Bold font (size 72). The corresponding print colours and their RGB codes were green (0, 128, 0), brown (165, 42, 42), hot pink (255, 105, 180), and white (255, 255, 255). These four words were non-cognates, that is, do not share phonological or orthographic features across languages, unlike several other colour word pairs (e.g., “blue/bleu” or “red/rouge”). Auditory stimuli consisted of the colour words (/vert/, /marron/, /rose/, /blanc/, French for green, brown, pink and white, respectively), spoken by a female speaker.

The manipulation allowed for 2 within-subject factors: Trial Type (“matching” condition that contained *All congruent* and *Sound-colour congruent* trials vs. “mismatching” condition that contained *Word-sound congruent*, *All incongruent,* and *Word-colour congruent* trials) and Language (French vs. English). In each experimental block, there were 25% matching (6.25% *All congruent*, 18.75% *Sound-colour congruent*) and 75% mismatching trials (18.5% *Word-sound* and *Word-colour congruent* trials, 37.5% *All incongruent)*. This was because each combination of colour word distracter, print colour, and sound were presented equally often to avoid contingency biases (i.e., learning of regularities between stimuli; Schmidt et al., 2007; see also, Lorentz et al., 2016).[1] This does mean that mismatching responses were more frequent than matching responses. However, it is important to note that all of the key comparisons are within response type. That is, we conducted one analysis for matching responses and another analysis for mismatching responses, as previously suggested (Goldfarb & Henik, 2006). This way, even if participants had a learned strategic tendency to prepare the "mismatching" response, this bias cannot impact "matching" responses. No systematic biases were produced in our statistical tests, as two trial types were analysed separately (i.e., none of our comparisons involve comparing a trial with a "matching" response to a "mismatching" response. In total, there were 3 larger experimental blocks of 128 trials each (in total 384 trials), presented randomly without replacement. This main phase of the experiment was preceded by a practice block. The practice block consisted of 32 trials, with the colour words replaced with the stimulus "xxxx".

Audiovisual Stimulus Combination

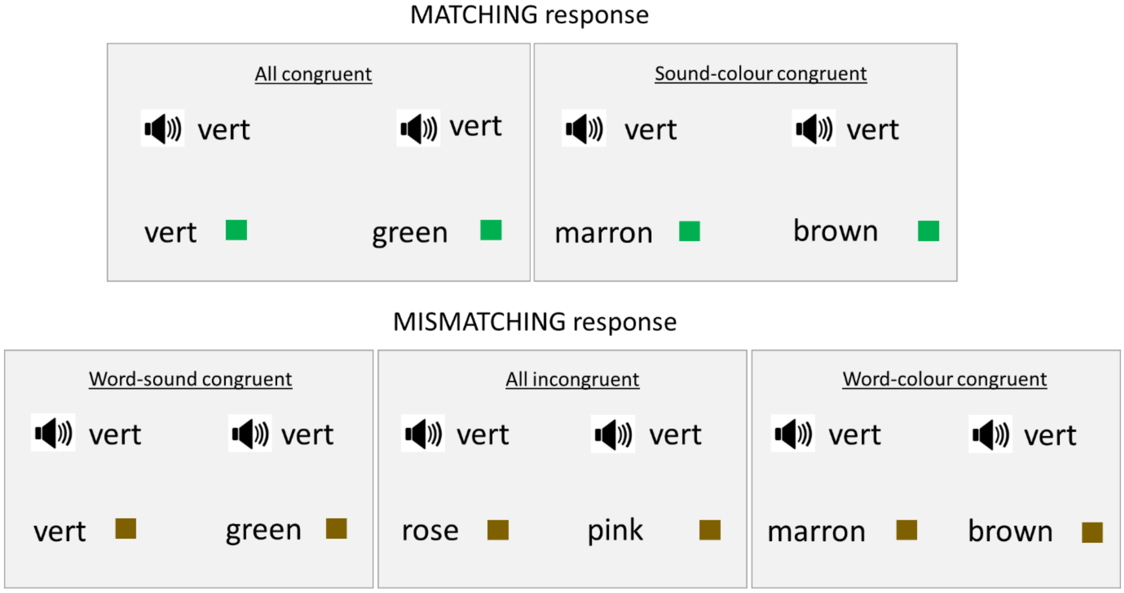
A total of 128 audiovisual stimulus combinations were created from the eight visual stimuli ("vert", "marron", "rose", "blanc", "green", "brown", "pink", "white"), four font colours (green, brown, pink, and white) and four auditory stimuli ("vert", "marron", "rose", "blanc"). These combinations were grouped into 5 conditions, varying by the congruence or incongruence between spoken word meaning, font colour, and written word meaning. In two conditions, the font colour and spoken colour word (task-relevant comparison) were congruent and thus required a "matching" response. These conditions were: 1) *All congruent*, and 2) *Sound-colour congruent*. In the other three conditions, the font colour and spoken colour word were incongruent and thus required a "mismatching" response. These conditions

[1] In a standard Stroop task, the proportion of congruent trials is often increased, sometimes merely to have the same number of congruent and incongruent trials (e.g., 1:1 congruent:incongruent in a four-choice task) or to increase control demands (e.g., 3:1 congruent:incongruent in Blumenfeld & Marian, 2014). However, this is suboptimal as regularities are introduced between distracting and target stimuli, meaning that congruency effects are confounded by contingency learning effects.

were: 3) *All incongruent*, 4) *Word-sound congruent*, and 5) *Word-colour congruent*. All of these five conditions applied for both distracter languages. These conditions are presented in Figure 3.

Figure 3

All trial types across two distracter languages (French and English)



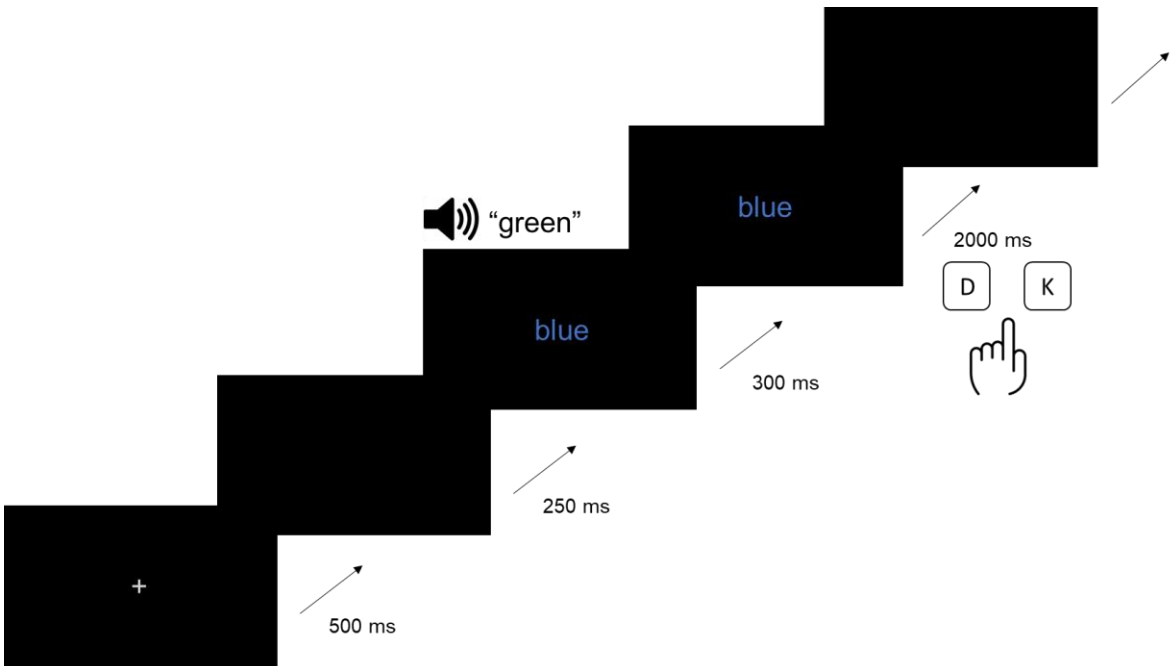
Note. All trial types have two equivalents; one with a French distracter (on the left) and one with an English distracter (on the right). Colour patches represent the ink colour in each trial.

Procedure

After completing the survey questions, the main experiment began. Stimuli were presented on a black (0, 0, 0) screen. On each trial, participants were first presented with a fixation "+" in grey (128, 128, 128) for 500 ms. This was followed by blank screen presented for 250 ms. Then the coloured distracter appeared on the screen until a response was registered or 2000 ms elapsed. The coloured distracter was presented simultaneously with the auditory stimulus. Responses could be provided only after 300 ms from the stimulus onset. This is due to the programming of the experiment. On each trial, an initial event plays the audio and presents the visual stimuli, which is then followed by a second event with only the stimulus and where responses are recorded. This was also done because the task required a comparison of the auditory stimulus with the print colour. Thus, a response before the auditory stimulus has been played is inevitably an anticipatory response that would be best excluded anyway. The next trial began after a 750-ms blank screen. The timeline of each trial is visualized in Figure 4. If the participant made an error or failed to respond in time, then the message “Erreur” (“Error”) or “Trop lent” (“Too slow”), respectively, appeared in red (255, 0, 0) for 1000 ms before the next trial. In both experiments, participants were explicitly instructed to respond as quickly and as accurately as possible and avoid reading a distracter since it represents a task-irrelevant dimension. The “matching” key had to be pressed for trials in which the spoken colour word and the font colour matched, and the “mismatching” key for trials in which the spoken colour word and the font colour mismatched.

Figure 4

Timeline of an experimental trial



Results

We used French and English words in this experiment to compare a highly-fluent L1 with a low-fluency L2. In [removed for review], French is normally the native language and English is typically learned later in life and not to a very high level of mastery. To assure that this was actually the case for our sample, we first analysed average language metric scores[2], which are presented in Table 1. All participants seemed to sufficiently fit our language criteria, as they were native French speakers who acquired the language early in life. Importantly, French was ranked as the first language in terms of dominance and order of acquisition by all participants. The percentage of French use revealed that participants had been using French almost exclusively in their everyday lives. In contrast, English was learned much later as a foreign language in primary schools. Participants were only moderately proficient in English, as shown by LexTale score and their self-rated English knowledge level. Although they studied English for a considerable amount of time (almost 9 years) and declared being able to speak and read English fluently (approximately at the age of 15), their objective proficiency level is rather low.

Table 1

Mean French and English language scores and standard errors (in brackets)

	M	*SE*
LexTale		
Years English	8.94 years	0.332
English level	3 (1-5)	0.158
Score	65.82 (0-100)	1.312

[2] The vast majority (33/34; 1 empty) of participants indicated French as their first language in order of dominance and in order of acquisition. One participant ranked Turkish as the first language in both dominance and acquisition, but further inspection of provided responses revealed that this participant had started acquiring French early enough and thus was therefore not excluded from the sample. As a second language in order of dominance and acquisition, participants rated English, followed by Spanish, Arabic, Creole, and Portuguese. The most frequently indicated third language in both dominance and acquisition were Spanish, German, English, Italian, Arabic, and Portuguese. All the participants correctly translated English words “green”, “brown”, “pink” and “white”.

LEAP-Q		
Dominance French	1	0
Dominance English	2.26	0.056
Order French	1	0
Order English	2.19	0.052
French Use (%)	4.97 (1-5)	0.029
English Use (%)	1.73 (1-5)	0.160
French		
Acquisition	1.10 years	0.183
Fluent	3.03 years	0.228
Reading	5.54 years	0.147
Fluent Read	6.79 years	0.198
English		
Acquisition	9.85 years	0.351
Fluent	15.41 years	0.344
Reading	12.42 years	0.386
Fluent Read	14.75 years	0.404

Data Analysis

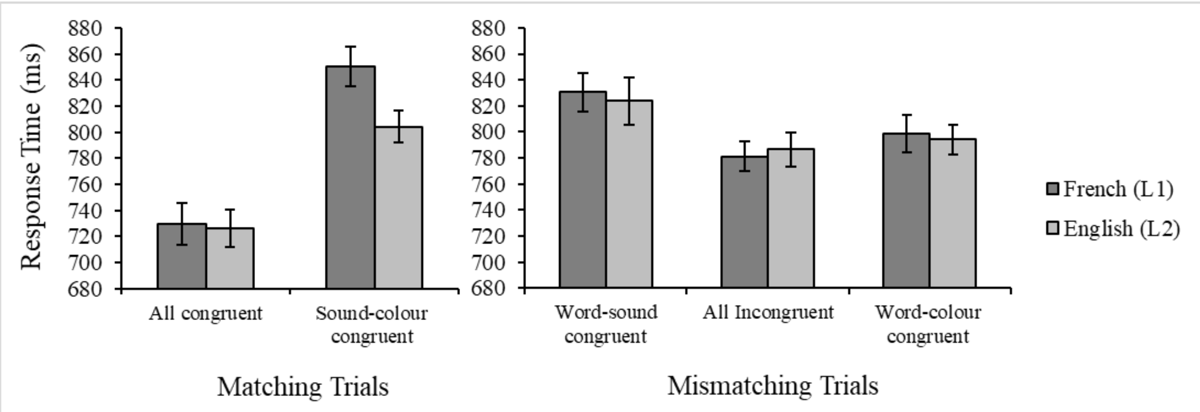
The mean correct response times (i.e., made during the 2000 ms response window) and mean percentage error were analysed. Response times were not trimmed (pre-planned analyses). However, we note that the direction and significance of all effects did not change in subsequent analyses with an Interquartile range (IQR) trim method, unless otherwise noted. No participants were excluded from the sample, as their individual accuracy rate was 86.35% or above. The congruency variable had different levels for “matching” and “mismatching” responses, and matching and mismatching trial types were analysed separately. One shared factor was a Distracter Language, with two levels: French (L1) and English (L2). Because the congruency variable had different levels for the “matching” and “mismatching” responses and because there are no relevant comparisons to make between the matching and mismatching trial types, two separate repeated measure analyses of variance with two within-subject factors were conducted. In the “matching” condition, 2 levels were analysed (*All congruent* and *Sound-colour congruent*), while in the “mismatching” condition, 3 levels were analysed (*Word-sound congruent*, *All incongruent* and *Word-colour congruent*).

Response time (RT)

Response times were recorded in milliseconds as the time elapsed from stimulus onset to key press. A total of 5.98% trials were excluded from the analyses (5.77% incorrect and .21% time out responses). Only RTs for correct responses in “matching” and “mismatching” conditions were analysed and illustrated in Figure 5.

Figure 5

Mean response times with standard errors for “matching” and “mismatching” trials



Matching trials

There was a main effect of Trial Type; $F(1,33) = 209.609$, $MSE = 1606.534$, $\eta_p^2 = .864$, $BF_{10} > 1000$, $p < .001$. Responses on *Sound-colour congruent* trials ($M = 827$, $SE = 13.30$) were slower than responses on *All congruent* trials ($M = 728$, $SE = 13.93$). The significant main effect of Language was observed, $F(1,33) = 11.638$, $MSE = 1797.765$, $\eta_p^2 = .260$, $BF_{10} = 1.124$, $p = .001$, with slower responses in French condition ($M = 790$, $SE = 14.71$) relative to English condition ($M = 765$, $SE = 12.53$). The interaction between Trial Type and Language was also significant, $F(1,33) = 9.272$, $MSE = 1649.944$, $\eta_p^2 = .219$, $BF_{10} = 11.021$, $p < .01$. There was no difference in response speed between French ($M = 729$, $SE = 16.06$) and English ($M = 726$, $SE = 14.45$) *All congruent* trials, $t(33) = .286$, $M_{diff} = 3$, $BF_{10} = .191$, $BF_{01} = 5.236$, $p = .776$. However, responses were significantly slower on French ($M =$

850, *SE* = 15.13) *Sound-colour congruent* trials relative to English *Sound-colour congruent* (*M* = 804, *SE* = 12.14) trials; *t*(33) = 6.847, M_{diff} = 46, BF_{10} > 1000, *p* < .001.

Mismatching trials

The main effect of Trial Type was observed, *F*(2,66) = 36.205, *MSE* = 926.505, η_p^2 = .523, BF_{10} > 1000, *p* < .001. Responses on *Word-sound congruent* (*M* = 827, *SE* = 15.79) trials were significantly slower than responses on *All incongruent* (*M* = 784, *SE* = 12.01) trials, *t*(33) = 7.156, M_{diff} = 43, BF_{10} > 1000, *p* < .001 and *Word-colour congruent* (*M* = 796, *SE* = 12.44) trials, *t*(33) = 5.085, M_{diff} = 31, BF_{10} > 1000, *p* < .001. Responses on *Word-colour congruent* trials were slower relative to responses on *All incongruent* trials, *t*(33) = 4.167, M_{diff} = 12, BF_{10} = 129.88, *p* < .001. There was no main effect of Language[3], *F*(1,33) = .278, *MSE* = 727.161, η_p^2 = .008, BF_{10} = .161, BF_{01} = 6.211, *p* = .602, indicating that there is no difference in response latencies between French and English trials. The interaction between Trial Type and Language was also not significant, *F*(2,66) = .664, *MSE* = 1031.101, η_p^2 = .02, BF_{10} = .179, BF_{01} = 5.586, p = .518.

Percentage error

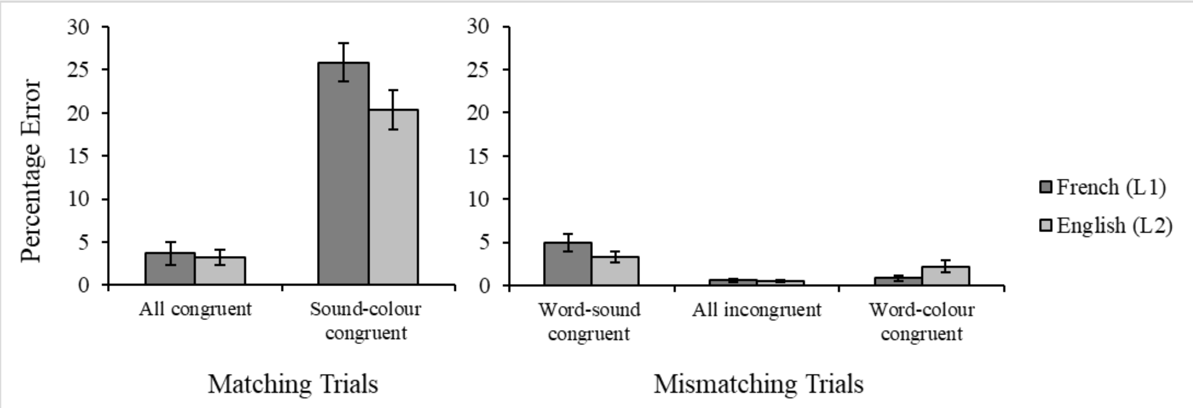
The mean percentage error data for all trial types and languages are presented in Figure 6.

---

[3] After trimming 512 outliers using the IQR method, the main effect of Language reached significance; F(1,33) = 6.243, *MSE* = 581.77, η_p^2 = .16, *p* = .02 for response times in Mismatching trials. Trials with French distracters (*M* = 781; *SE* = 10.54) were responded to slower than trials with English distracters (*M* = 773, *SE* = 10.17).

Figure 6

Mean percentage error with standard error for "matching" and "mismatching" trials



Matching trials

There was a main effect of Trial Type, $F(1,33) = 113.835$, $MSE = 115.229$, $\eta_p^2 = .775$, $BF_{10} > 1000$, $p < .001$, indicating that participants made significantly more errors on *Sound-colour congruent* ($M = 23.07$, $SE = 2.08$) than on *All congruent* trials ($M = 3.43$, $SE = .89$). The main effect of Language was observed, $F(1,33) = 8.034$, $MSE = 37.752$, $\eta_p^2 = .196$, $BF_{10} = .391$, $BF_{01} = 2.557$, $p = .01$, with higher percentage errors on French ($M = 14.75$, $SE = 1.43$) than on English trials ($M = 11.76$, $SE = 1.39$). The interaction between Trial Type and Language was marginally significant, $F(1,33) = 4.272$, $MSE = 49.6$, $\eta_p^2 = .115$, $BF_{10} = .987$, $BF_{01} = 1.013$, $p = .05$. There was no significant difference in percentage error between French ($M = 3.68$, $SE = 1.37$) and English ($M = 3.19$, $SE = .86$) *All congruent* trials, $t(33) = .338$, $M_{diff} = .49$, $BF_{10} = .194$, $BF_{01} = 5.155$, $p = .737$. However, participants made significantly more errors on French ($M = 25.81$, $SE = 2.23$) than on English ($M = 20.33$, $SE = 2.29$) *Sound-colour congruent* trials, $t(33) = 3.144$, $M_{diff} = 5.483$, $BF_{10} = 10.617$, $p < .01$, similar to the response time data.

Mismatching trials

There was a main effect of Trial Type, $F(2,66) = 19.381$, $MSE = 11.884$, $BF_{10} > 1000$, $\eta_p^2 = .37$, $p < .001$. That is, participants made significantly more mistakes in *Word-sound congruent* ($M = 4.095$, $SE = .69$) relative to *All incongruent* ($M = .532$, $SE = .118$) trials, $t(33) = 5.524$, $M_{diff} = 3.563$, $BF_{10} > 1000$, $p < .001$), and *Word-colour congruent* ($M = 1.513$, $SE = .456$) trials, $t(33) = 3.826$, $M_{diff} = 2.583$, $BF_{10} = 54.49$, $p = .001$. The percentage error was larger in the *Word-colour congruent* than in the *All incongruent* condition, $t(33) = 2.329$, $M_{diff} = .98$, $BF_{10} = 1.93$, $p < .05$. No significant main effect of Language was observed, $F(1,33) = .102$, $MSE = 6.423$, $\eta_p^2 = .003$, $BF_{10} = .154$, $BF_{01} = 6.493$, $p = .752$. The interaction between Trial Type and Language was significant, $F(2,66) = 5.112$, $MSE = 7.647$, $\eta_p^2 = .134$, $BF_{10} = 3.078$, $p = .01$. There were no significant differences in percentage errors between French and English *Word-sound congruent* trials, $t(33) = 1.788$, $M_{diff} = 1.645$, $BF_{10} = .766$, $BF_{01} = 1.305$, $p = .083$ and *All incongruent* trials, $t(33) = .397$, $M_{diff} = .08$, $BF_{10} = .198$, $BF_{01} = 5.05$, $p = .694$. However, participants made significantly more errors on English than French *Word-colour congruent* trials, $t(33) = 2.223$, $M_{diff} = 1.386$, $BF_{10} = 1.587$, $p < .05$.

Correlations

As a supplementary analysis, we assessed the level to which language metric variables correlate with different types of trials with both French (L1) and English (L2) colour words used in the Stroop matching task. These analyses were purely exploratory and did not reveal any clear or significant results. However, we present these data in the Appendix for the interested reader.

Discussion

Experiment 1 had two aims: 1) compare the magnitude of between-language and within-language interference, and 2) investigate the source of interference in a bilingual

Stroop matching task with intermixed French (L1) and English (L2) colour word distracters. Within-language interference was larger than between-language interference, but only for *Sound-colour congruent* trials, with no significant difference between French and English word pairs across other trial types. That is, when a spoken word (e.g., “vert”, French for green) matched the ink colour of the written distracter, the French incongruent distracters (e.g., “marron”, French for brown printed in green) were responded to slower and less accurately than English incongruent distracters (e.g., “brown” in green). It is plausible that French written distracters lead to a strong task-irrelevant comparison (i.e., written word-spoken word) that impairs performance on a task-relevant comparison (i.e., ink colour-spoken word). *Sound-colour congruent* trials also had significantly higher percentage errors relative to all other trial types. This is probably due to the fact that both task-irrelevant comparisons activate the “mismatching” response in contrast to task-relevant comparison which activates the “matching” response. However, the observed pattern of results for both French and English “matching” trials clearly correspond to the assumptions of both stimulus and response conflict, with faster responses on *All congruent* relative to *Sound-colour congruent* trials.

Theoretically more interesting are the results for the mismatching trial types. Responses on *Word-sound congruent* trials were significantly slower and more error prone relative to *All incongruent* and *Word-colour congruent* trials (Bornstein, 2015). That is, both incongruent French (e.g., “vert” in brown) and English (e.g., “green” in brown) distracters slowed down responding when the word distracter corresponded to the auditory stimulus (e.g., hear “vert”). This contrasts with the results of Goldfarb and Henik (2006), who found the slowest “mismatching” responses for congruent distracters (i.e., *Word-colour congruent* trials). Interestingly, response latencies were almost identical in French and English condition, suggesting that responding to the spoken L1 word is equally affected by a written L1 word (i.e., both spoken and written words are identical) and an L2 word (i.e., spoken and written words are not identical, but represent the same colour concept, e.g., “vert” and “green”).

The responses were the fastest in *All incongruent* condition, which confirms the assumptions of the response conflict account. This also aligns with the findings on behavioural data of Caldas and colleagues (2012) and Goldfarb and Henik (2006), thus confirming a role of response conflict in the Stroop matching task. In contrast, the semantic conflict account should have predicted that these trials would be the *slowest*, because the word, colour, and auditory stimulus are all incongruent with each other.

Experiment 2

Experiment 2 conceptually replicates Experiment 1 with one important modification. In particular, instead of the colour words used in Experiment 1, participants were presented with French and English colour associates. A complication with the matching task is that the predictions for the stimulus and response conflict accounts for mismatching trials are exactly in opposition. The response conflict account predicts that *All incongruent* trials should be the fastest of the three “mismatching” trial types (as observed), whereas the semantic conflict account predicts that they should be the slowest. Note that the predictions of both semantic and response conflict accounts for colour associates are identical to the predictions for colour words, already visualised in Figure 1. If both types of conflict exist, then it might be that the (larger) response conflict effect is concealing a (relatively smaller) semantic conflict effect. Therefore, one way to “reveal” the true effect of semantic conflict (assuming there is one, of course), would be to eliminate the response conflict. According to some, colour associates produce semantic conflict (e.g., (Glaser & Glaser, 1989; Schmidt & Cheesman, 2005), but not response conflict. If this logic is correct, it remains plausible that semantic conflict will be observed for colour associates. Although probably smaller, semantic conflict might emerge due to strong conceptual links between colour associates and their corresponding colour words. For example, on a French *Sound-colour congruent* trial (e.g., see “ciel”, French for sky, printed in green, hear “vert”, French for green), a distracter “ciel”, associated with blue, should no longer interfere (or very little) with a relevant task comparison (i.e., “green”-“green”), simply because it does not belong to the same semantic category as a spoken word. Experiment 2 was therefore designed to further explore the role of semantic conflict that was possibly masked by response conflict in Experiment 1. Another question of interest concerns the distracter language. According to some models of bilingual memory, L2 words do not have strong direct access to semantics (Kroll & Stewart, 1994). Thus, while semantic conflict might be observed for L1 words, these models would predict the absence of a semantic conflict effect for L2 words.

Method

Participants

A total of 33 (25 women) [removed for review] undergraduates (M_{age} = 20; *SD* = 3.43) voluntarily participated in the experiment. The sample size was determined in the same way as in Experiment 1. All the selection criteria were identical to Experiment 1. Students who already participated in Experiment 1 were not allowed to participate in Experiment 2. Their average language background scores (mean age and standard errors) are presented in Table 2 (see Results section).

Apparatus and materials, design, and procedure

Experiment 2 was identical in all aspects to Experiment 1, with the following exceptions. First, colour words were replaced by colour associates in French (L1) and English (L2), which correspond to “blue”, “green”, “red”, and “yellow”, respectively (i.e., “ciel”/“sky”, “herbe”/“grass”, “sang”/“blood”, and “citron”/“lemon”). These words were non-cognates with a mean word length of 4.75 for French associates and 4.5 for English associates. The colour associates from both languages were chosen based on: 1) their strong association with a corresponding colour word (Nelson et al., 1998; Wilson et al., 1988) and 2) their similarity in word length. Second, in line with used colour associates, spoken words were “bleu” (blue), “vert” (green), “rouge” (red), and “jaune” (yellow). All trial timings were identical to Experiment 1.

Results

Average language metric scores[4] are presented in Table 2. As in Experiment 1, participants started acquiring French at early age (as it is a native language), while English was learned as a first foreign language in schools (starting at around 10 years old), but again, not to a very high level of mastery. Similar to Experiment 1, participants had rather low objective English proficiency, as shown by the LexTale score, as well as low self-estimated English level. All participants seemed to sufficiently fit our language dominance criteria.

Table 2

Mean French and English language scores for and standard errors (in brackets)

	M	*SE*
LexTale		
Years English	10 years	0.484
English level	2.79 (1-5)	0.155
Score	67.91 (0-100)	1.531
LEAP-Q		
Dominance French	1	0
Dominance English	2.1	0.049
Order French	1	0
Order English	2.06	0.045
French Use (%)	4.85 (1-5)	0.063
English Use (%)	1.82 (1-5)	0.147

[4] All the participants (33/33) indicated French as their first language in order of dominance and in order of acquisition. As a second language in order of dominance and in order of acquisition, participants mostly indicated English, followed by German, Spanish, and Vietnamese. The most frequent third language in dominance and order of acquisition was Spanish, followed by German, English, Italian, and Polish. The majority of participants correctly translated “sky” (31/33), “blood” (32/33), and “lemon” (32/33). However, only half of them correctly translated “grass” (17/33).

French		
Acquisition	0.59 years	0.179
Fluent	3.17 years	0.287
Reading	5.44 years	0.162
Fluent Read	6.61 years	0.252
English		
Acquisition	9.62 years	0.510
Fluent	15.2 years	0.458
Reading	11.39 years	0.486
Fluent Read	15.48 years	0.543

Data Analysis

As in Experiment 1, the mean correct response times and mean percentage error [5]were analysed. No participants were excluded from the sample, their individual accuracy rate across the experiment was 89.84% or above. Two separate ANOVAs (one for Matching trials and one for Mismatching trials) were conducted for both response times and percentage errors.

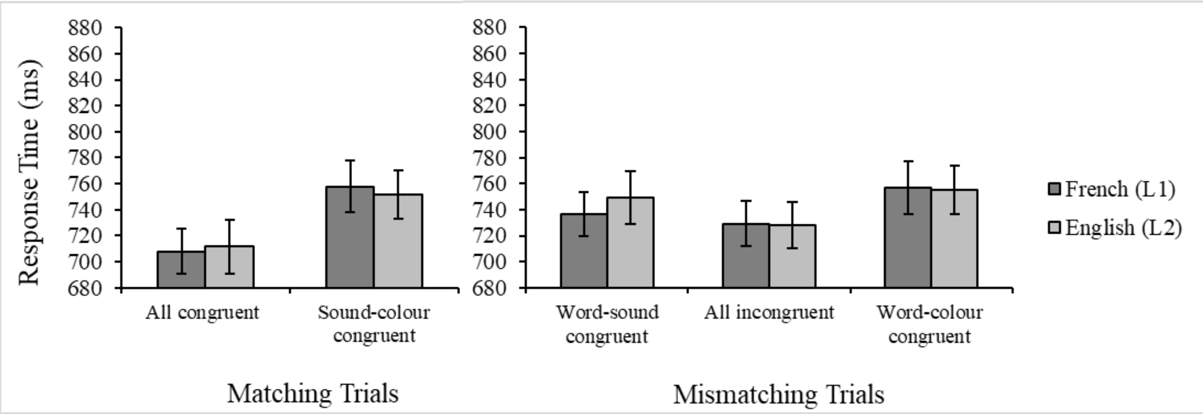
Response time (RT)

A total of 5.03% trials were excluded from the analyses (4.65% incorrect and .38% time out responses). Only RTs for correct responses in Matching and Mismatching conditions were analysed and illustrated in Figure 7.

[5] We note that subsequent analyses revealed that response time and error results were largely similar for all four words. It seems plausible that while recall (i.e., translation) was rather low for “grass”, participants were probably able to recognize the English word during the task.

Figure 7

Mean response times with standard errors for "matching" and "mismatching" trials



Matching trials

There was a main effect of Trial Type, $F(1,32) = 32.467$, $MSE = 2043.097$, $\eta_p^2 = .504$, $BF_{10} > 1000$, $p < .001$, suggesting that responses on *Sound-colour congruent* trials ($M = 754$, $SE = 18.71$) were significantly slower than responses on *All Congruent* trials ($M = 710$, $SE = 18.24$). However, there was no main effect of Language, $F(1,32) = .041$, $MSE = 1280.291$, $\eta_p^2 = .001$, $BF_{10} = .182$, $BF_{01} = 5.494$, $p = .840$, indicating no overall difference in response speed to French and English word trials. The interaction between Trial Type and Language was also not significant, $F(1,32) = .364$, $MSE = 2425.755$, $\eta_p^2 = .011$, $BF_{10} = .348$, $BF_{01} = 2.873$, $p = .550$.

Mismatching trials

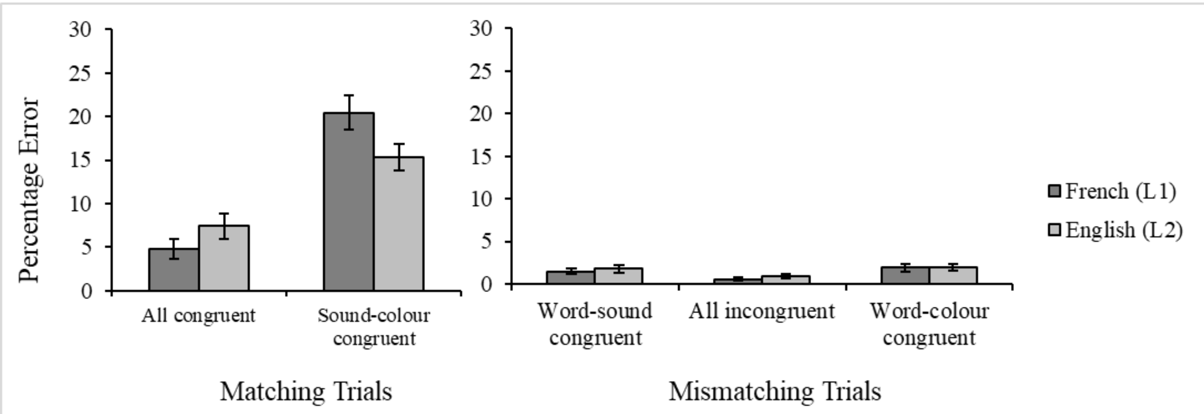
The main effect of Trial Type was observed, $F(2,64) = 21.143$, $MSE = 589.472$, $\eta_p^2 = .398$, $BF_{10} > 1000$, $p < .001$. *Word-colour congruent* trials ($M = 756$, $SE = 18.87$) were responded to slower than *All incongruent* ($M = 729$, $SE = 17.46$) trials, $t(32) = 6.293$, $M_{diff} = 27$, $BF_{10} > 1000$, $p < .001$, and *Word-sound congruent* ($M = 743$, $SE = 17.99$) trials, $t(32) = 3.004$, $M_{diff} = 13$, $BF_{10} = 7.70$, $p = .01$. Responses were slower on *Word-sound congruent* relative to *All incongruent* trials, $t(32) = 3.663$, $M_{diff} = 14$, $BF_{10} = 35.69$, $p < .01$. There was no main effect of Language, $F(1,32) = .581$, $MSE = 882.089$, $\eta_p^2 = .018$, $BF_{10} = .193$, $BF_{01} = 5.181$, $p = .451$, suggesting no overall difference in response speed between French and English word trials. The interaction between Trial Type and Language was also not significant, $F(2,64) = 1.073$, $MSE = 1043.801$, $\eta_p^2 = .032$, $BF_{10} = .25$, $BF_{01} = 4$, $p = .348$.

Percentage error

The mean percentage error data for all trial types and languages are presented in Figure 8.

Figure 8

Mean percentage errors with standard errors for "matching" and "mismatching" trials



Matching trials

There was a main effect of Trial Type, $F(1,32) = 77.71$, $MSE = 58.774$, $\eta_p^2 = .708$, $BF_{10} > 1000$, $p < .001$, suggesting that *Sound-colour congruent* trials ($M = 17.859$, $SE = 1.498$) were significantly more error-prone relative to *All congruent* trials ($M = 6.095$, $SE = .969$). No main effect of Language was observed, $F(1,32) = 1.32$, $MSE = 38.6$, $\eta_p^2 = .04$, $BF_{10} = .233$, $BF_{01} = 4.292$, $p = .259$, suggesting no overall difference in percentage error between French and English word trials. An interaction between Trial Type and Language was significant, $F(1,32) = 7.839$, $MSE = 61.967$, $\eta_p^2 = .197$, $BF_{10} = 12.331$, $p = .01$. That is, there was no difference in percentage error between French (*M* = 4.798, *SE* = 1.149) and English (*M* = 7.392, *SE* = 1.422) *All congruent* trials, $t(32) = 1.516$, $M_{diff} = 2.594$, $BF_{10} = .525$, $BF_{01} = 1.905$, $p = .139$. However, participants made significantly more errors on French (*M* = 20.399, *SE* = 1.966) than on English (*M* = 15.32, *SE* = 1.486) *Sound-colour congruent* trials, $t(32) = 2.854$, $M_{diff} = 5.079$, $BF_{10} = 5.56$, $p = .01$.

Mismatching trials

A main effect of Trial Type was significant, $F(2,64) = 7.53$, $MSE = 3.182$, $\eta_p^2 = .19$, $BF_{10} = 34.428$, $p = .001$. Participants made significantly more errors on *Word-colour congruent* trials (*M* = 1.91, *SE* = .32) relative to *All incongruent* (*M* = .75, *SE* = .18) trials, $t(32)= 4.06$, $M_{diff} = 1.16$, $BF_{10} = 96.42$, $p < .001$. There was no difference in percentage error between *Word-colour congruent* and *Word-sound congruent* (*M* = 1.61, *SE* = .33) trials, $t(32) = .873$, $MEAN_{diff} = .30$, $BF_{10} = .26$, $BF_{01} = 3.85$, $p = .389$. Participants made more errors on *Word-sound congruent* relative to *All incongruent* trials, $t(32) = 2.86$, $M_{diff} = .862$, $BF_{10} = 5.63$, $p < .05$. There was no significant main effect of Language, $F(1,32) = 1.179$, $MSE = 2.931$, $\eta_p^2 = .035$, $BF_{10} = .243$, $BF_{01} = 4.115$, $p = .286$. An interaction between Trial Type and Language was also not significant, $F(2,64) = .154$, $MSE = 3.435$, $\eta_p^2 = .005$, $BF_{10} = .105$, $BF_{01} = 9.524$, $p = .858$.

Correlations

As in Experiment 1, we assessed the level to which language metric variables correlate with different trial types with both French (L1) and English (L2) colour associates used in the Stroop matching task. Similar to Experiment 1, there were no significant correlations. However, we present these data in the Appendix.

Discussion

Experiment 2 aimed to 1) compare the magnitude of between-language and within-language interference produced by French (L1) and English (L2) colour associates, and 2) investigate the source of this interference. In line with the predictions of both semantic and response conflict accounts, *Sound-colour congruent* trials are responded to slower and with more errors relative to *All congruent* trials. Interestingly, a lack of interaction suggests that participants were equally fast in responding to French and English distracters. This contrasts the assumption of larger within-language (i.e., produced by French distracters) relative to between-language (i.e., produced by English distracters) interference.

Concerning the “mismatching” trials, *Word-colour congruent* trials were responded to slower than *Word-sound congruent* and *All incongruent* trials, suggesting that congruent colour associates (e.g., “ciel” in blue or “sky” in blue) interfere with “mismatching” responses, as observed by Goldfarb and Henik (2016) and Caldas et al. (2012) with colour words. It is plausible that participants take additional time to process the congruency of the to-be-ignored written colour associates, which slows down responding. Interestingly, almost equal response times were observed with both French and English distracters, suggesting that first and second language distracters might be processed in a similar way.

Finally, responses were again the fastest on *All incongruent* trials, which aligns with the assumption of the response conflict account. That is, even for colour associate distracters, participants perform all three task comparisons, which suggest the same, “mismatching” response alternative. Thus, contrary to expectations, the use of colour associates did not eliminate response conflict, allowing us to observe a potential true (but small) semantic conflict effect. Instead, colour associates (like colour words) seemingly produced response conflict.

General Discussion

The present study aimed to explore the effects of bilingual colour word and colour associate distracters on matching stimuli presented in auditory (i.e., spoken word) and visual (i.e., ink colour) formats. In Experiment 1, participants were presented with either congruent or incongruent colour words in French (L1) and English (L2), accompanied with a spoken French colour word. Experiment 2 followed the same logic, but French and English colour associates appeared as distracters. In both experiments, participants were explicitly instructed to ignore the colour word and to respond based on whether ink colour and spoken word match or mismatch. This manipulation allowed comparisons between two matching trial types (*All congruent* and *Sound-colour congruent*) and three mismatching trial types (*Word-sound congruent*, *All incongruent*, and *Word-colour congruent*).

The first question of interest concerns the language of distracters. Since only French colour words were used as spoken stimuli, French distracters should produce within-language interference, whereas English distracters should produce between-language interference. As already discussed in the Introduction, previous findings suggest that within-language interference is usually larger than between-language interference (Fang et al., 1981; Kiyak, 1982; MacLeod, 1991). We observed this pattern with the matching trial types, where there was evidence for a larger congruency effect for L1 than L2. No language differences were found for mismatching trial types, however. This makes the findings similar to those expected for more balanced bilinguals. It is important to note that participants were tested only on a small set of words (i.e., colour words), which are often learned in early phases of second language learning. It would be interesting to test these finding with less balanced bilinguals or by using a larger set of distracting words, which might reveal clearer differences between L1 and L2 items. Future work may also make use of mixed modelling of individual-trial response times, as traditional methods of data analysis do not always account for individual differences across bilingual participants (Privitera et al., 2023). Alternatively, L2 words might possess a strong link with their corresponding conceptual representations, similarly to L1 words (Šaban & Schmidt, 2021; Schmidt et al., 2018). As discussed in the Introduction, L2 words could possess strong semantic connections, lexical connections, or a combination of both. Therefore, the nature of L2 connections and their strength towards lexical and semantic representations should help elucidate the similarities/differences observed in patterns for both L1 and L2 words.

However, it seems that the difference in magnitudes of within- and between-language interference is even smaller for colour associates (Experiment 2) relative to colour words (Experiment 1). That is, overall response times were faster for colour associates than for colour words (Schmidt & Cheesman, 2005). Moreover, no difference was observed between French and English trials, thus suggesting that the first and second language colour associates seem to interfere less with L1 spoken colour words relative to colour word distracters. This can be due to the fact that colour associates, although semantically related to colour words, do not correspond to the spoken colour words. This finding thus revealed that the meaning of the written distracter, either from L1 or L2, cannot be completely ignored, resulting in a decrease of the response speed within which ink colour and spoken words were judged as "matching" or "mismatching". This interference produced by written distracters seems to increase proportionally with its similarity to the spoken word. That is, in both experiments, spoken words were French colour words. Responses were generally slower in Experiment 1 when the same set of French colour words was used as distracters. That is, written, to-be-ignored colour word distracters also served as potential targets. In contrast, responses were faster in Experiment 2 when colour associates were used as distracters. Although these colour associates were semantically related to spoken colour words, they were not targets. This aligns with the assumptions of the *response set membership* account (Klein, 1964; Risko et al., 2006), which refers to a larger interference for words (e.g., distracters) that are potential targets (e.g., or a to-be-attended stimulus dimension, such as a spoken word in the Stroop matching task). This has been confirmed with both colour words and colour associates (Klein, 1964; Risko et al., 2006; Schmidt & Cheesman, 2005; Sharma & McKenna, 1998) in the literature and in the present series of experiments.

A second question of interest is the source of interference produced in the Stroop matching task. The semantic conflict account suggests that responses will be the slowest on trials in which task dimensions activate multiple colour concepts. For instance, larger interference is expected on trials in which two contrasting colour representations are simultaneously activated (e.g., *Sound-colour congruent* trials) relative to trials in which only one colour representation is activated (e.g., *All congruent* trials). In contrast, the response conflict account focuses on task comparisons and assumes that responses will be slowest on trials in which task-relevant and task-irrelevant comparisons suggest different responses. That is, responses should be faster on trials in which all three task comparisons suggest the same response option (e.g., “match” or “mismatch”, for *All congruent* and *All incongruent* trials, respectively), relative to those trials in which one comparison activates one response option, whereas two other comparisons activate contrasting response option (e.g., on *Word-sound congruent* or *Word-colour congruent* trials). The interplay between semantic and response conflict is also possible. For instance, these two conflict effects might be in opposition in the matching task. That is, the larger response conflict is “masking” the smaller semantic conflict. One way to measure the true effect of semantic conflict would be to eliminate the response conflict. To do so, colour associates (which are assumed to produce semantic conflict exclusively) were used as alternative to colour words in Experiment 2.

As expected, the response times were slower for incongruent trials (i.e., *Sound-colour congruent*) relative to congruent trials (i.e., *All congruent*) with “matching” response.

However, previous findings suggest that the response times are slower for congruent relative to incongruent trials with “mismatching” responses (Bornstein, 2015; Caldas et al., 2012; Goldfarb & Henik, 2006). That is, *Word-colour congruent* trials (e.g., “green” in green, hear “pink”) are assumed to be responded to slower than *All incongruent* (e.g., “green” in brown, hear “pink”) and *Word-sound congruent* (e.g., “green” in brown, hear “green”). This has been replicated in Experiment 2 using colour associates, when *Word-colour congruent* trials (e.g., “sky” in blue, hear “green”) produced the slowest response latencies as compared to other two types of trial. However, this pattern was not observed in Experiment 1 which made use of colour words. In Experiment 1, the responses where slowest on *Word-sound congruent* trials (e.g., “green” in brown, hear “green”). That is, instead of focusing on congruency of the written stimuli exclusively, as suggested by previous studies, participants tend to compare a written, to-be-ignored stimulus, with a spoken word, thus engaging a task-irrelevant comparison.

Navon (1985) introduced the notion of outcome-conflict to reflect a state where the output of one task modifies (and potentially interferes) a variable that is relevant to the performance of a concurrent task (Navon, 1985; Navon & Miller, 1987). In this conceptualization, performance in the Stroop matching task is determined by a conflict of outcomes between three separate dimensions, each one resulting in either a “matching” or “mismatching” outcome. It is possible that outcome-conflicts occurred whenever the relevant matching task and the two mistakenly performed matching tasks produced conflicting outcomes (i.e., “matching” vs. “mismatching”). Interference effects were large and significant only in conditions that featured such a conflict. For instance, outcome-conflict does not predict any interference in *All congruent* and *All incongruent* conditions because all three comparisons between colour representations indicate the same response, “matching” and “mismatching”, respectively. According to this account, when one irrelevant matching outcome conflicted with the response (e.g., on *Word-sound congruent* and *Word-colour congruent* trials, when a correct response was “mismatch”, and two irrelevant comparisons suggest “match” and “mismatch”), the interference should be smaller than on trials in which both irrelevant outcomes conflicted with the response (e.g., on *Sound-colour congruent* trials when a correct response was “match”, but both irrelevant comparisons suggest “mismatch”). In sum, as the number of outcome-conflicts becomes larger, performance is more prone to errors. Our results align with this: the percentage error was extremely high in the *Sound-colour congruent* condition relative to remaining four trial types (in both Experiment 1 and Experiment 2). Consequently, to achieve higher accuracy, participants probably focus on serial processing of separate comparisons, which in turn might have produced additional response delays. This is also observable in the present results, with *Sound-colour congruent* trials being slower relative to all other trial types.

The present findings also align with the confluence model proposed by Eviatar and colleagues (1994) based on their findings from a visual matching task. According to this model, in matching tasks, all stimulus dimensions are processed automatically and simultaneously regardless of task relevance. This processing and an interference produced by the outputs between all task dimensions precede response selection. In the present study, visual and auditory stimuli were processed until their representations could be compared. The “matching” or “mismatching” among the outputs of these comparisons determined the response speed and the likelihood of selecting the correct response alternative. This interpretation is similar to the one proposed by Navon’s (1985) outcome-conflict account. However, this confluence model is more specifically oriented toward matching tasks and more explicit regarding the processing stage to which interference is attributed (Eviatar et al., 1994).

The present findings with colour word distracters (Experiment 1) align with behavioural data of Caldas and colleagues (2012) and those of Goldfarb and Henik (2006), providing an additional support for the response conflict account. Interestingly, we observed response conflict effect even with colour associates, which we assumed (incorrectly) would eliminate the response conflict component. However, the electrophysiological data of Caldas and colleagues (2012) supported a semantic conflict account. This data showed that conflict related brain activity, as indicated by a greater frontal negativity (N450), was not observed for a "mismatching" condition that featured conflicting irrelevant "matching" output. Rather, N450 amplitude was greater in *Word-colour congruent* and *All incongruent* conditions than in the *Word-sound congruent* condition. This discrepancy between behavioural and electrophysiological data suggests that interference produced in the Stroop matching task could be due to contributions of both semantic and response conflict. It is plausible that the role of semantic conflict in explaining the Stroop matching interference could be evidenced exclusively by using more subtle measures, such as electrophysiology. Another possibility is that there still might be a semantic conflict effect observable in behavioural studies, however, it is still being masked by response conflict.

The present results clearly indicate a role for response conflict in the Stroop matching task, for colour words and colour associates and in the first and less-fluent second language. However, the role of semantic conflict is less clear. As highlighted in this manuscript, one peculiarity of the matching task is that it can only provide evidence for either response conflict or semantic conflict, but not both, as the two are pitted against each other. As such, it is not currently clear whether semantic conflict was absent in our studies, or rather merely smaller than (and therefore concealed by) response conflict. Future research could help answering these inquiries. Indeed, as indicated in the Introduction, one of the goals of the present manuscript was to assess some competing models of bilingual memory. According to certain models, stimulus conflict should only occur for L1 words in early language learners, but not for L2 words, whereas other models suggest that stimulus conflict should occur for both. Given the absence of stimulus conflict in the present task, even for L1 words, we were unable to assess such competing models with the current data. In sum, despite the fact that response conflict plays an important role in the interference produced in the Stroop matching task, this does not discard the possibility that some other, non-response (i.e., semantic) conflict also contributes to this effect, which remains a focus of debate (Caldas et al., 2020; Dittrich & Stahl, 2017; Green et al., 2016; Luo, 1999).

Conclusion

The present experiments explored how different types of first and second language words influence audio-visual matching performance. The findings suggest that, regardless of the distracting language (L1 vs. L2), responses were the fastest on trials in which task comparisons activate fewer response alternatives, supporting the assumption of the response conflict account. That is, performance is faster when no competition between response alternatives occurs. The present work serves as a good starting point in understanding how simultaneous audio-visual processing affects response selection across languages and word types.

References

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., & Ferrand, L. (2014). Automaticity of word reading: Evidence from the semantic Stroop paradigm. *Current Directions in Psychological Science*, *23*(5), 343–348. https://doi.org/10.1177/0963721414540169

Besner, D., Stolz, J. A., & Boutilier, C. (1997). The stroop effect and the myth of automaticity. *Psychonomic Bulletin & Review*, *4*(2), 221–225. https://doi.org/10.3758/BF03209396

Blumenfeld, H. K., & Marian, V. (2014). Cognitive control in bilinguals: Advantages in Stimulus–Stimulus inhibition. *Bilingualism: Language and Cognition*, *17*(3), 610–629. https://doi.org/10.1017/S1366728913000564

Bornstein, I. S. (2015). *Behavioural measures of interference and facilitation in an audiovisual colour-word Stroop matching task*. The University of British Columbia.

Caldas, A. L., Machado-Pinheiro, W., Daneyko, O., & Riggio, L. (2020). The Stroop-matching task as a tool to study the correspondence effect using images of graspable and non-graspable objects. *Psychological Research*, *84*(7), 1815–1828. https://doi.org/10.1007/s00426-019-01191-5

Caldas, A. L., Machado-Pinheiro, W., Souza, L. B., Motta-Ribeiro, G. C., & David, I. A. (2012). The Stroop matching task presents conflict at both the response and nonresponse levels: An event-related potential and electromyography study: Stroop matching task and conflict. *Psychophysiology*, *49*(9), 1215–1224. https://doi.org/10.1111/j.1469-8986.2012.01407.x

Dittrich, K., & Stahl, C. (2017). Two distinct patterns of interference in between-attribute Stroop matching tasks. *Attention, Perception, & Psychophysics*, *79*(2), 563–581. https://doi.org/10.3758/s13414-016-1253-x

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

Dyer, F. N. (1973). Interference and facilitation for color naming with separate bilateral presentations of the word and color. *Journal of Experimental Psychology*, *99*(3), 314–317. https://doi.org/10.1037/h0035245

Egeth, H. E., Blecker, D. L., & Kamlet, A. S. (1969). Verbal interference in a perceptual comparison task. *Perception & Psychophysics*, *6*(6), 355–356. https://doi.org/10.3758/BF03212790

Eviatar, Z., Zaidel, E., & Wickens, T. (1994). Nominal and physical decision criteria insame-different judgments. *Perception & Psychophysics*, *56*(1), 62–72. https://doi.org/10.3758/BF03211691

Fang, S.-P., Tzeng, O. J. L., & Alva, L. (1981). Intralanguage vs. Interlanguage Stroop effects in two types of writing systems. *Memory & Cognition*, *9*(6), 609–617. https://doi.org/10.3758/BF03202355

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Ferrand, L., & Augustinova, M. (2014). Differential effects of viewing positions on standard versus semantic Stroop interference. *Psychonomic Bulletin & Review*, *21*(2), 425–431. https://doi.org/10.3758/s13423-013-0507-z

Flowers, J. H. (1975). "Sensory" interference in a word-color matching task. *Perception & Psychophysics*, *18*(1), 37–43. https://doi.org/10.3758/BF03199364

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

Goldfarb, L., & Henik, A. (2006). New Data Analysis of the Stroop Matching Task Calls for a Reevaluation of Theory. *Psychological Science*, *17*(2), 96–100. https://doi.org/10.1111/j.1467-9280.2006.01670.x

Goldfarb, L., & Henik, A. (2007). Evidence for task conflict in the Stroop effect. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(5), 1170–1176. https://doi.org/10.1037/0096-1523.33.5.1170

Green, M. L., Locker, L., Boyer, T. W., & Sturz, B. R. (2016). Stroop-like interference in a match-to-sample task: Further evidence for semantic competition? *Learning and Motivation*, *56*, 53–64. https://doi.org/10.1016/j.lmot.2016.09.003

Hamers, J. F., & Lambert, W. E. (1972). Bilingual interdependencies in auditory perception. *Journal of Verbal Learning and Verbal Behavior*, *11*, 303–310.

Kiyak, H. A. (1982). Interlingual interference in naming color words. *Journal of Cross-Cultural Psychology*, *13*(1), 125–135.

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

Lorentz, E., McKibben, T., Ekstrand, C., Gould, L., Anton, K., & Borowsky, R. (2016). Disentangling Genuine Semantic Stroop Effects in Reading from Contingency Effects: On the Need for Two Neutral Baselines. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00386

Luo, C. R. (1999). Semantic competition as the basis of Stroop interference: Evidence from color-word matching tasks. *Psychological Science*, *10*(1), 35–40. https://doi.org/10.1111/1467-9280.00103

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

Mägiste, E. (1982). Automaticity and interference in bilinguals. *Psychological Research*, *44*(1), 29–43. https://doi.org/10.1007/BF00308553

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)

Morton, J. (1969). Categories of interference: Verbal mediation and conflict in card sorting. *British Journal of Psychology*, *60*(3), 329–346. https://doi.org/10.1111/j.2044-8295.1969.tb01204.x

Navon, D. (1985). Attention division or attention sharing? In *Attention and performance XI* (pp. 133–146). Erlbaum.

Navon, D., & Miller, J. (1987). Role of outcome conflict in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 435–448.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. http://www.usf.edu/FreeAssociation/

Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In *Information Processing and cognition: The Loyola Symposium* (pp. 55–85). Hillsdale, NJ: Erlbaum.

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

Privitera, A. J., Momenian, M., & Weekes, B. S. (2023). Modeling the bilingual advantage: Do results differ between analysis methods? *Ampersand*, *11*, 100134. https://doi.org/10.1016/j.amper.2023.100134

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

Šaban, I., & Schmidt, J. R. (2021). Stimulus and response conflict from a second language: Stroop interference in weakly-bilingual and recently-trained languages. *Acta Psychologica*, *218*, 103360. https://doi.org/10.1016/j.actpsy.2021.103360

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., 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., 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

Seymour, P. H. K. (1977). Conceptual encoding and locus of the Stroop effect. *Quarterly Journal of Experimental Psychology*, *29*(2), 245–265.

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

Simon, J. R., & Berbaum, K. (1988). Effect of irrelevant information on retrieval time for relevant information. *Acta Psychologica*, *67*(1), 33–57. https://doi.org/10.1016/0001-6918(88)90023-6

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.

Tanaka, J. W., & Presnell, L. M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics*, *61*(6), 1140–1153. https://doi.org/10.3758/BF03207619

Treisman, A., & Fearnley, S. (1969). The Stroop Test: Selective Attention to Colours and Words. *Nature*, *222*(5192), 437–439. https://doi.org/10.1038/222437a0

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

Virzi, R. A., & Egeth, H. E. (1985). Toward a translational model of Stroop interference. *Memory & Cognition*, *13*(4), 304–319. https://doi.org/10.3758/BF03202499

Wilson, M., Kiss, G., & Armstrong, C. (1988). *University of Oxford, EAT: the Edinburgh associative corpus*. http://hdl.handle.net/20.500.12024/1251

Declaration

Conflict of interest

The authors have no conflicts of interest to declare.

Replication package

Research materials

All research materials, including participant recruitment material, questionnaire, task instructions and debriefing form are available at https://osf.io/48q2p/?view_only=3c4cdec3f832446984291fc5f22f6392 under the section "Research materials"

Data

Data is available at https://osf.io/48q2p/?view_only=3c4cdec3f832446984291fc5f22f6392 under the section "Data".

Analysis code

Instructions and code required to reproduce all analyses are available at https://osf.io/48q2p/?view_only=3c4cdec3f832446984291fc5f22f6392 under the section "Analysis code".

Appendix

Table A1 presents the non-parametric rank-based Spearman's correlation coefficients between the behavioural measures (i.e., response times and error rates) and language metric scores for Experiment 1. We observed that only percentage error, but not response speed, correlated with certain language metric variables (e.g., age of development of English reading skills or percentage of English exposure). Note however that after applying a Holm-Bonferroni correction for multiple comparisons, none of the correlations were significant at *a* = .05, so these correlations should be interpreted with caution.

Table A1

Correlations between behavioural and language metric scores in Experiment 1

	French (L1)										English (L2)									
	Matching						Mismatching				Matching						Mismatching			
	All congruent		Sound-colour congruent		Word-sound congruent		All incongruent		Word-colour congruent		All congruent		Sound-colour congruent		Word-sound congruent		All incongruent		Word-colour congruent	
	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR
LexTALE	-.186	.022	-.211	.157	-.090	-.054	-.093	-.054	-.107	-.121	-.248	-.126	-.123	.157	-.086	-.054	-.132	-.015	-.132	.212
English Level	-.128	-.296	-.027	-.077	.113	.294	.098	*-.385*	.053	.047	.021	-.002	-.029	.034	.103	.077	.047	.161	.099	-.116
Years English	.032	.145	.043	-.035	.171	.033	.183	-.081	.139	.166	-.117	-.113	.197	-.128	.318	-.005	.179	-.098	.176	-.194
% French Exposure	.027	.096	.098	-.098	.062	-.193	.009	.110	.062	*-.390*	-.044	-.227	.027	.196	-.009	-.028	.115	.111	.062	-.311
% English Exposure	.032	-.214	-.086	*-.430*	.025	-.068	.123	-.318	.014	-.125	.046	-.032	-.074	-.185	.152	.066	.056	-.242	.086	-.193
FRENCH																				
Acquisition	.093	.136	.090	-.146	-.061	*-.339*	.020	-.218	-.082	-.297	.234	.058	.068	-.139	-.101	**-.456**	.005	-.123	.127	-.238
Fluent	-.202	.023	-.090	.304	-.068	.080	-.202	.069	-.240	-.083	-.150	.073	-.158	.160	-.136	-.004	-.119	.069	-.032	.070
Reading	-.019	-.009	.044	.138	.049	.241	.026	.131	-.057	-.134	.068	-.153	.068	.240	.089	.059	.060	.151	.184	.067
Fluent Reading	.062	.307	.029	.206	-.015	.034	-.071	.057	-.080	*-.364*	.148	-.027	-.025	.327	-.004	-.017	.026	-.053	.069	-.089
ENGLISH																				
Acquisition	-.113	-.063	-.088	.017	.052	.076	-.134	.059	-.070	.009	.009	-.073	-.161	-.173	-.134	.213	-.084	.032	-.134	.129
Fluent	.139	.213	.135	.180	-.082	.051	.067	-.028	-.025	-.105	.075	.014	.242	.188	.082	-.097	.019	-.189	.068	.012
Reading	-.021	**.470**	-.141	.081	.006	.061	-.109	.200	-.083	.021	-.240	-.114	-.193	.114	-.024	*.393*	-.069	-.090	-.146	.185
Fluent Reading	-.079	.327	-.187	.128	-.026	.156	-.057	.052	-.073	.052	-.203	-.087	-.198	-.024	.077	*.388*	-.023	-.256	-.109	.179

Note. Italic = $p < .05$, Bold = $p < .01$; no tests were significant after Holm-Bonferroni correction.

Table A2 presents the same correlation for the Experiment 2 data. As in Experiment 1, none of the correlations were significant at a = .05 after applying the Holm-Bonferroni correction for multiple comparisons. As such, the following should be interpreted with caution. We observed that the response speed for all trial types (both French and English) were negatively correlated with the age of reading in French. That is, the earlier participants started reading in French, the slower their responses were. This seems reasonable because reading is often considered as an automatic skill (Augustinova & Ferrand, 2014) acquired early in life. However, in this task, participants were explicitly instructed to avoid reading a distracter since it represents a task-irrelevant dimension and impairs matching/mismatching responses.

Table A2

Correlations between behavioural and language metric scores in Experiment 2

	French (L1)										English (L2)									
	Matching				Mismatching						Matching				Mismatching					
	All congruent		Sound-colour congruent		Word-sound congruent		All incongruent		Word-colour congruent		All congruent		Sound-colour congruent		Word-sound congruent		All incongruent		Word-colour congruent	
	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR	RT	ERR
LexTALE	.293	.016	.109	.077	.112	.101	.054	-.285	.111	-.109	.202	.172	.183	.067	.166	.048	.084	.024	-.001	.043
English Level	.121	.323	.039	-.008	.026	.082	-.040	*-.369*	.033	-.237	.096	.141	.022	.057	.018	-.143	-.039	.098	.011	-.050
Years English	.161	.181	.095	-.188	.198	-.116	.246	-.117	.135	.065	.140	.189	.200	-.325	.112	-.282	.173	.099	.172	*-.357*
% French Exposure	.124	-.213	.275	.286	.302	.228	.186	.256	.266	.094	*.435*	.327	.284	.254	.266	.214	.266	.232	.337	.171
% English Exposure	-.141	-.094	.052	-.129	-.043	.237	-.039	*-.415*	-.010	-.203	-.009	.048	-.011	-.235	-.078	-.153	-.078	.147	-.075	.116
FRENCH																				
Acquisition	-.141	.202	-.072	-.061	-.117	-.136	-.123	.107	-.271	.110	-.095	.041	-.228	-.097	-.142	-.078	-.135	-.045	-.061	-.092
Fluent	.055	-.017	.248	.084	.262	.066	.286	.152	.238	.262	.203	.209	.120	.199	.196	.059	.176	.089	.226	.280
Reading	*-.411*	.056	**-.497**	.030	*-.454*	.100	**-.505**	.150	**-.467**	-.064	**-.465**	-.159	*-.422*	.165	**-.464**	-.063	**-.482**	.210	*-.445*	.136
Fluent Reading	-.116	.255	-.189	-.078	-.158	-.196	-.087	.187	-.108	.040	-.196	-.106	-.150	-.003	-.111	-.239	-.127	.133	-.097	-.189
ENGLISH																				
Acquisition	.149	.123	.148	-.001	.110	.043	.085	-.057	.101	-.101	.283	.118	.155	.092	.163	.085	.129	-.069	.067	.224
Fluent	.329	-.120	.107	-.178	.147	.031	.159	**-.545**	.216	-.313	.080	-.234	.262	-.164	.221	-.130	.099	-.263	-.023	.112
Reading	.040	.039	.085	-.281	.020	-.225	.065	-.312	.116	*-.408*	.036	-.284	.135	-.238	.092	-.328	.128	-.030	.041	-.017
Fluent Reading	.004	-.176	.061	-.237	-.012	-.070	.077	**-.520**	.029	-.335	-.137	*-.382*	.064	-.293	.008	-.230	.013	-.238	-.092	.088

Note. Italic = $p < .05$, Bold = $p < .01$; no tests were significant after Holm-Bonferroni correction.