Interference in Dutch-French Bilinguals:

Stimulus and Response Conflict in Intra- and Interlingual Stroop

James R. Schmidt,[1] Robert J. Hartsuiker,[2] and Jan De Houwer[1]

[1]Department of Experimental Clinical and Health Psychology, Ghent University

[2]Department of Experimental Psychology, Ghent University

Word count: 4832

Author Notes

Abstract

In the present manuscript, we investigate the source of congruency effects in a group of Dutch-French bilinguals. In particular, participants performed a colour-identification Stroop task, in which both (first language) Dutch and (second language) French distracting colour words were presented in colours. The typical finding is impaired responding when the word and colour are incongruent (e.g., "red" in blue) relative to congruent (e.g., "red" in red). This congruency effect is observed for both first and second language distracting colour words. The current experiment used a 2-to-1 keypress mapping manipulation, which allows one to separate stimulus conflict (i.e., conflict between word and colour meanings) and response conflict (i.e., conflict between potential responses). For both the first and second language, both stimulus and response conflict were observed. These results suggest that second language words influence semantic and response processing similarly to first language words, rather than having diminished semantic and/or response influences.

Keywords: bilingualism; Stroop effect; stimulus conflict; response conflict; semantics; response selection

Introduction

Considerable research has focused on bilingual cognition (for reviews, see Bialystok, Craik, & Luk, 2012; Koda, 1996; Werker & Byers-Heinlein, 2008). One particular focus is on how second language words influence cognitive processing in a qualitatively different (or not) way than first language words. In the present report, we explore the source of crosstalk between languages within a bilingual version of the colour-word Stroop paradigm. In particular, we investigate the extent to which foreign words might have diminished impact on semantic identification and/or response decision processes, or to what extent first and second language words might influence cognitive processing in a similar fashion.

In the Stroop task (Stroop, 1935; for a review, see MacLeod, 1991), participants are tasked with the goal of identifying the print colour of a colour word, while ignoring the meaning of the word itself (e.g., say "green" to the word "red" printed in green). The *congruency (or Stroop) effect* is the observation that participants are typically slower and less accurate to *incongruent* trials (e.g., the word "green" printed in yellow), where the meaning of the word and colour mismatch, relative to *congruent* trials (e.g., "green" in green), where the meaning of the word and colour match.

One question of interest in the Stroop literature is the origin of the conflict. That is, what produces the conflict between a colour word and an incongruent colour? On the one hand, conflict could occur between the *meaning* of the word and of the colour (e.g., lexical-semantic representations), which we term here *stimulus conflict* (Glaser & Glaser, 1989; Mackinnon, Geiselman, & Woodward, 1985; Stirling, 1979). An alternative possibility is that the response engendered by the word and the response engendered by the colour compete for selection, which we term here *response conflict* (Klein, 1964; Posner & Presti, 1987). As will be discussed shortly, the general consensus is that both stimulus and response conflict contribute to the standard Stroop effect (Augustinova & Ferrand, 2014; Augustinova, Silvert,

Ferrand, Llorca, & Flaudias, 2015; Ferrand & Augustinova, 2014).

One clear line of evidence for both stimulus and response contributions to the Stroop effect comes from 2-to-1 mapping experiments. For instance, De Houwer (2003; see also, A. T. Chen, Bailey, Tiernan, & West, 2011, 2004; Hasshim & Parris, 2015; Jongen & Jonkman, 2008; van Veen & Carter, 2005) presented participants with a 2-to-1 Stroop task, in which participants responded to two colours for each key. For instance, a given participant might have been instructed to respond to the colours blue and yellow with the left key and the colours green and red with the right key, as illustrated in Figure 1. This produces three conditions, rather than just two. First, there are *identity* trials (e.g., "blue" in blue), which are typical congruent trials in which the word matches the colour. The response to the word, by extension, also matches the response to the colour. Second are *same response* trials (e.g., "blue" in yellow), which are incongruent in meaning (i.e., blue and yellow are different colours) but mapped to the same response key (i.e., the responses for blue and yellow are both the left key). A difference between identity and same response trials therefore indicates stimulus conflict, and not response conflict. Third are *different response* trials (e.g., "blue" in green), in which the word mismatches the meaning of the colour (as in same response trials), but the assigned responses also mismatch (i.e., the response keys for blue and green are different). Any difference between same and different response trials thus indicates response conflict. Evidence for both stimulus and response conflict was observed, with same response trials being slower than identity trials (stimulus conflict), but faster than different response trials (response conflict).
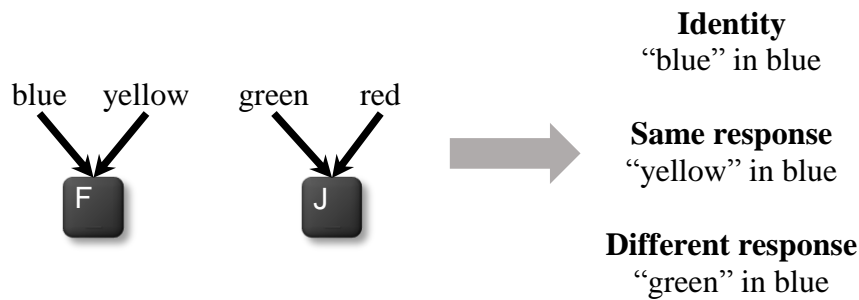
Identity
"blue" in blue

blue   yellow   green   red

F   J

Same response
"yellow" in blue

Different response
"green" in blue

Figure 1. Illustration of the 2-to-1 mapping procedure. In addition to stimulus-/response-
compatible identity trials, and stimulus-/response-incompatible different response
trails, there are also stimulus-incompatible but response-compatible same response
trials.

Not all stimulus types produce *both* stimulus and response conflict. For instance,

colour associates also produce a congruency effect, with incongruent colour associates (e.g.,

"sky" in red) impaired relative to congruent colour associates (e.g., "sky" in blue). Following

debate about whether this effect was due to stimulus or response conflict (Glaser & Glaser,

1989; Klein, 1964; Mackinnon et al., 1985; Posner & Presti, 1987; Stirling, 1979), Schmidt

and Cheesman (2005) used the 2-to-1 mapping procedure and found exclusively stimulus

conflict for colour associates (cf., Risko, Schmidt, & Besner, 2006). This was interpreted as

indicating that associates spread activation to related concepts in semantics, producing

semantic conflict with the target colour concept, but are not potent enough to indirectly bias a

potential response (e.g. "sky" facilitating "blue" strongly enough to retrieve a D-key response

linked to "blue," which also applies in a verbal task where "sky" is not a potential response).

Similarly, response conflict can be observed for distracting stimuli where stimulus conflict

would be impossible, such as in the Simon task (Simon, Craft, & Webster, 1973; Simon &

Rudell, 1967); there can be no stimulus conflict between a colour and a location as these

involve different stimulus dimensions. Thus, both stimulus and response conflict contribute to

performance in an intralingual (within language) Stroop procedure, though both types of

interference are not necessarily observable for all stimulus types. In sum, stimulus and

response conflict do not necessarily develop in parallel for any given association (see also,

Schmidt, Crump, Cheesman, & Besner, 2007; Schmidt & De Houwer, 2012).

The Stroop effect has also been used extensively to study interference between two languages (interlingual) in bilingual participants (Altarriba & Mathis, 1997; Atalay & Misirlisoy, 2012; H. C. Chen & Ho, 1986; Dalrymple-Alford, 1968; Dyer, 1971; La Heij et al., 1990; Mägiste, 1984, 1985; Preston & Lambert, 1969; Smith & Kirsner, 1982; Tzelgov, Henik, & Leiser, 1990). It is known from this research that a Stroop effect can be observed with both colour word distracters of the *first language (L1)* and of the *second language (L2)*. For instance, a native English speaker that also speaks French will be impaired by incongruent French colour words, in addition to incongruent English colour words. However, the standard finding is that the effect for L2 words is smaller than that for L1 words. For instance, the native English speaker described above will be more impaired by a trial such as "yellow" in green than by a trial such as "jaune" (French for "yellow") in green (for a review, see MacLeod, 1991).

Aside from the issue of the overall size of the congruency effect in native and foreign languages is the *source* of the conflict. The key question of the present manuscript is whether L2 distracting words engender both stimulus and response conflict or only one of the two. This can have important implications for theorizing about language cognition, as much debate centers on how language lexicons are connected to each other and to semantics. For instance, consider the account provided by Kroll and Stewart (1994), presented in Figure 2. According to this account, the L1 and L2 lexical representations for words are connected. For instance, "zwart" and "noir" (respectively, Dutch and French for "black") are connected in memory. In addition, both words are connected to a single semantic representation for black. It is additionally assumed in this model that: (a) L1 words are more strongly connected to semantic concepts (i.e., heavy overlearning) than L2 words, and (b) L2 words are more strongly connected to L1 words (i.e., learned as translations of the already well-known first
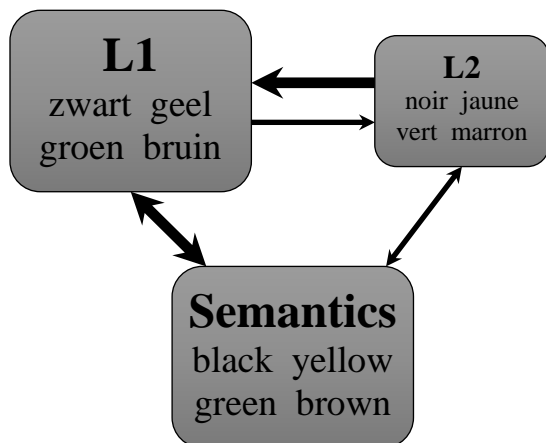
language).

Figure 2. Illustration of the Kroll and Stewart (1994) model of the connections between first (L1) and second (L2) lexicons and semantics. Note the asymmetries in the connections between lexicons and to semantics.

Of course, there are many alternative versions of the model presented in the figure (e.g., which involve different connection strengths between memory stores). However, the key question of the current report is *where* conflict occurs within the system: at the level of semantics, at the level of responses, or both. In that vein, the present work made use of the 2-to-1 mapping design[1] that included colour words from both the dominant L1 language (Dutch) and a non-dominant L2 language (French). Three possible results might occur. Of course, the first is that both stimulus and response conflict will be observed for L2 words, just as with L1 words. In this case, we would expect the same pattern of results for L2 French colour words (i.e., identity < same responses < different response).

However, conflict effects from L2 words are generally smaller than from L1 words, which might mean that one and/or both of the conflict components are decreased for L2 words. A second possibility, therefore, is that foreign colour words produce exclusively stimulus conflict. For instance, Altarriba and Mathis (1997) argued that foreign colour words

---

[1] It is relevant to point out that this design makes use of a keypress rather than verbal modality. Added differences between languages might be observable that are specific to a verbal response modality (e.g., conflict during articulation), which we discuss in the Discussion section.

are directly linked to semantics, even at early stages of learning a new language. According to

this view (which contrasts sharply with that presented in Figure 2), the foreign word "noir"

interferes with the semantic identification of yellow print colour (i.e., slower answer to

question "What colour is that?"). That is, presentation of the word "noir" will activate the

semantic representation for black, and the yellow print colour activates the semantic

representation for yellow, leading to conflict in semantics (cf., Schmidt, Cheesman, & Besner,

2013). However, depending on further assumptions one makes, "noir," according to this

account, will not interfere afterward when a response (keypress) is being selected. If so, then

L2 colour words act like colour associates to the L1 equivalents (e.g., "noir" as an associate to

"black"). Like with colour associates, the assumption would be that "noir" is unable to

retrieve the response linked to black, even though (unlike with colour associates) "noir" is a

direct lexical translation of "black." If this is the case, then we should expect a difference

between identity and same response trials for L2 colour words, but no difference between

same and different response trials.

An third possibility is that foreign colour words might produce exclusively response

conflict. For instance, it could be the case that "noir" interferes with what response needs to

be made (i.e., slower answer to the question "What key do I need to press?" or, in verbal

naming, "What do I need to verbalize?"). For this to occur, the foreign distracting word would

need to be able to automatically bias a response without interfering with identification of the

stimulus colour itself. For instance, it might be that stimulus conflict occurs exclusively

within semantics, to which L2 words are weakly connected (Kroll & Stewart, 1994), as

illustrated earlier in Figure 2. Instead, L2 words might be quickly and automatically translated

to their L1 lexical equivalents (e.g., because the words were *learned* as translations), allowing

an indirect biasing of responses via encoded lexical-response instruction memories.

As an added consideration, either or both of the stimulus and response conflict effects

for L2 colour words might depend on the *type* of colour word. *Cognates* are translation equivalents with similar spellings in both languages, such as "bleu" (blue) in French and "blauw" in Dutch, which are typically similar because of a shared etymology. Of course, for cognates there is a compatibility in pronunciation (in the case of a verbal task), in addition to orthographic (spelling) similarities between the words. *Non-cognates*, in contrast, are dissimilar, such as "jaune" (yellow) in French and "geel" in Dutch. For non-cognates, there is little or no overlap in pronunciation/spelling. Indeed, similarity between the words in the two languages does matter for the amount of interference observed (e.g., Dyer, 1971), with larger interference effects for cognates. Because effects for cognates might be due to processes that are less interesting for our present purposes (see Costa, Miozzo, & Caramazza, 1999), such as priming across languages due to spelling similarities (or even just first letter priming), we opted to use Dutch-French non-cognates (Dutch/French: zwart/noir [black], groen/vert [green], bruin/marron [brown], and geel/jaune [yellow]).

Method

Participants

Ninety-three Ghent University undergraduates (71 female, 22 male) participated in the study in exchange for €5. We aimed for a large sample relative to previous experiments with the 2-to-1 mapping procedure because we were testing for potential language differences of unknown size. The exact sample size was determined by the number of participants who signed up during the allotted testing time. On the recruitment website, we explicitly solicited participants who were native Dutch speakers with some familiarity with French. Language questionnaires (to be discussed shortly) were used to confirm the fit of participants with these criteria. Average language metric scores are presented in Table 1. All participants seemed to sufficiently fit our language dominance criteria, so no participants were removed from the

sample. To broadly characterize the sample, English is learned as the second foreign language in schools and practically everyone in Flanders develops near native-speaker proficiency in the language.[2] French, on the other hand, while learned earlier on in school, is not nearly as well developed in the Flemish population. This was borne out in the language metrics. While participants self-rated their French proficiency relatively moderately (6.05 on a 0-10 scale), their more objective French test scores (see description of the LEXTALE_FR later) were quite low. Indeed, even the highest scorers in the sample were nowhere near the range of scores for native speakers. Thus, our sample was generally familiar with French, but had only weak French skills. See the Appendix for more detailed information on the language demographics.

Table 1. Mean language scores with standard errors.

	Mean	SE
LEXTALE_FR		
Years French	7.93 years	0.144
French Level	6.05 (0-10)	0.137
Score	5.23*	0.911
LEAP-Q		
Dominance Dutch	1.02**	0.015
Dominance French	3.10**	0.061
Order Dutch	1.09**	0.039
Order French	2.29**	0.056
%Dutch Use	71.3%	1.68
%French Use	7.4%	0.97
Dutch		
Acquisition	1.1 years	0.19
Fluent	4.7 years	0.25
Reading	6.1 years	0.15
Fluent Read	8.1 years	0.21
French		
Acquisition	9.2 years	0.30
Fluent	13.1 years	0.37
Reading	11.2 years	0.22
Fluent Read	14.0 years	0.25

*1st percentile L1, 48th L2, **ranks from 1st up.

---

[2] For this reason, we also ensured that the French stimuli were non-cognates with English colour words.

Apparatus and materials

The main part of the experiment was programmed in E-Prime 2 (Psychology Software Tools; Pittsburgh, PA) and conducted on a standard PC. Responses were made with the "F" (left) and "J" (right) keys on an AZERTY keyboard with the two index fingers. Prior to the computer portion of the experiment, participants were also given a short pen-and-paper survey to fill out. This included the LEXTALE_FR (Brysbaert, 2013) with Dutch-language instructions. In this test, participants are presented with a list of 84 French-looking words, only about 2/3 of which are actual French words (e.g., "église"), whereas the remaining 1/3 are not (e.g., "metter"). The participants are informed to select the words that they are fairly certain are actual French words. Correct "hits" are rewarded with one point, and incorrect "false alarms" are penalized by two points. Random guessing will therefore produce a score of around zero, with higher scores for better hit-to-false alarm ratios. The questionnaire also asks for gender, native language, years of French training in school, and a self-rating of French knowledge ranging from 0 (almost none) to 10 (perfect). Appended to this were a subset of questions from the Dutch for Belgium version of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007). In particular, the first three questions were retained, which asked, respectively, for a list of languages in order of dominance, a list of languages in order of acquisition, and the percentage with which the participant used each of their languages in the recent period. Also retained from the LEAP-Q were two boxes, one for Dutch and one for French, asking for the age the participant began acquiring the language, became fluent in the language, began learning to read in the language, and became fluent in reading the language. The purpose of these questionnaires was primarily to assure participants had the correct language dominance, but we also consider correlations of these metrics with the observed congruency effects. Finally, as an addition to these two questionnaires, participants were asked to give the French

translations of the four Dutch colour words used in the experiment. This was to get a general idea of how familiar the stimuli were to participants (see Appendix for a summary) and to make sure the participants knew the correct translation of each of the colours.

Design

During the main part of the experiment, participants were presented with the Dutch and French colour words for "black," "green," "brown," and "yellow" (Dutch/French: "zwart/noir," "groen/vert," "bruin/marron," and "geel/jaune," respectively). Notably, these four words are non-cognates, unlike several other colour words (e.g., "blauw/bleu" [blue], "rood/rouge" [red], etc.). The corresponding print colours were black (0,0,0), green (0,128,0), brown (139,69,19), and yellow (255,215,0), corresponding to "black," "green," "saddlebrown," and "gold" in the standard E-Prime colour palette. For each participant, two colours were mapped to the left key (e.g., black and yellow), and two to the right key (e.g., brown and green). Which colours were mapped to which keys and in which combinations was fully counterbalanced across participants (i.e., six factorial combinations) on the basis of participant number. These manipulations allow for two within factors. The first is distracter language (Dutch vs. French). The second is congruency: *identity* when the word and colour match, *same response* when the mismatching word and colour are mapped to the same key, and *different response* when the mismatching word and colour are mapped to different keys. In total there were three larger blocks of trials, separated by a five second pause. Each of the larger blocks consisted of two smaller sub-blocks. In each sub-block, each of the eight words were presented once each in all four colours (32 trials), selected randomly without replacement. Thus, there were 192 experimental trials total across the six sub-blocks. The main phase of the experiment was also preceded by a practice block. Similar to the main phase, the practice block had two sub-blocks of 32 trials each. However, the colour words were replaced with the stimulus "xxxx," randomly presented eight times in each of the colours

per sub-block.

Procedure

After completing the survey questions on pen and paper (see above), the main experiment began on the computer. Stimuli were presented on a white (255,255,255) screen in 18 pt., bold Courier New font. On each trial, participants were first presented with a fixation "+" in grey (128,128,128) for 250 ms. This was followed by a blank screen for 250 ms. Next, the coloured word was presented until a response was registered or 2000 ms elapsed. The next trial began immediately following a correct response. If the participant made an error or failed to respond in 2000 ms, then the message "Fout" ("False/Error") or "Te Traag" ("Too slow"), respectively, appeared in red (255,0,0) for 1000 ms before the next trial.

<div align="center">

Results

</div>

Both mean correct response time and percentage error data were assessed for the computer portion of the task. For response times, only correct responses were considered, but no other trims were made. For error percentages, trials in which participants failed to respond before the trial ended were excluded (0.3% of trials).

Response times

The correct response time data are presented in Figure 3. To analyse response times, we conducted a language (Dutch vs. French) by congruency (identity vs. same response vs. different response) within-subjects repeated measures ANOVA. The main effect of congruency was significant, $F(2,184) = 22.840$, $MSE = 1486$, $p < .001$, $\eta_p^2 = .20$. However, there was no main effect of language, $F(1,92) = 0.183$, $MSE = 1525$, $p = .670$, $\eta_p^2 < .01$, indicating no overall difference in response speed to Dutch and French word trials. Most importantly, the interaction between language and congruency was not significant, $F(2,184) = 0.703$, $MSE = 1239$, $p = .497$, $\eta_p^2 < .01$. Despite this lack of an interaction, we conducted

planned comparisons on each language separately. For Dutch colour words, there was both a

significant stimulus conflict effect (same response – identity), $t(92) = 2.409$, $SE_{diff} = 6$, $p =$

.018, $\eta^2 = .06$, and response conflict effect (different response – same response), $t(92) =$

3.348, $SE_{diff} = 5$, $p = .001$, $\eta^2 = .11$. Similarly for French colour words, both stimulus conflict,

$t(92) = 2.322$, $SE_{diff} = 5$, $p = .022$, $\eta^2 = .06$, and response conflict, $t(92) = 2.182$, $SE_{diff} = 5$, $p =$

.032, $\eta^2 = .05$, were observed. There was no evidence for any differences in the magnitude of

the stimulus conflict, $t(92) = 0.319$, $SE_{diff} = 8$, $p = .750$, $\eta^2 < .01$, or response conflict effects,

$t(92) = 0.903$, $SE_{diff} = 7$, $p = .369$, $\eta^2 < .01$, across languages, though at least a numerical trend

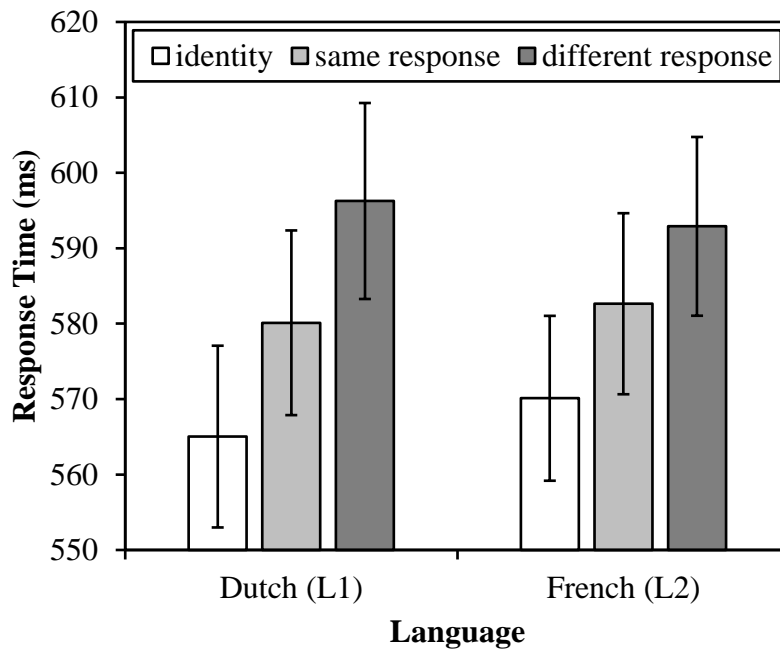for smaller effects in French (particularly for response conflict).

Figure 3. Response times with standard errors for Dutch and French colour words.

Percentage errors

The percentage error data are presented in Figure 4. We again conducted a language

(Dutch vs. French) by congruency (identity vs. same response vs. different response) within-

subjects repeated measures ANOVA. The main effect of congruency was significant,

$F(2,184) = 7.195$, $MSE = 24$, $p < .001$, $\eta_p^2 = .07$. However, there was no main effect of

language, $F(1,92) = 0.036$, $MSE = 14$, $p = .850$, $\eta_p^2 < .01$, indicating no overall difference in

error rates between Dutch and French word trials. Most importantly, the interaction between language and congruency was not significant, $F(2,184) = 2.285$, $MSE = 28$, $p = .105$, $\eta_p^2 = .02$, albeit with a trend toward a larger effect for Dutch. Again, we conducted planned comparisons on each language separately. For Dutch colour words, there was no stimulus conflict effect (same response – identity), $t(92) = 2.344$, $SE_{diff} = .8$, $p = .281$, $\eta^2 = .06$, but there was a significant response conflict effect (different response – same response), $t(92) = 4.266$, $SE_{diff} = .7$, $p = .001$, $\eta^2 = .17$. For French colour words, there was neither a stimulus conflict effect, $t(92) = 2.632$, $SE_{diff} = .8$, $p = .243$, $\eta^2 = .07$, nor a response conflict effect, $t(92) = 1.309$, $SE_{diff} = .7$, $p = .194$, $\eta^2 = .02$. There was no evidence for any difference in the magnitude of the stimulus conflict effect across languages, $t(92) = 0.132$, $SE_{diff} = 1.1$, $p = .895$, $\eta^2 < .01$. However, the response conflict effect was marginally larger in Dutch than in French, $t(92) = 1.978$, $SE_{diff} = 1.0$, $p = .051$, $\eta^2 = .04$. Similar to the response times, then, there were some hints of larger effects for L1 colour words, particularly for response conflict. However, despite the relatively large sample size, these differences were not sufficiently robust.
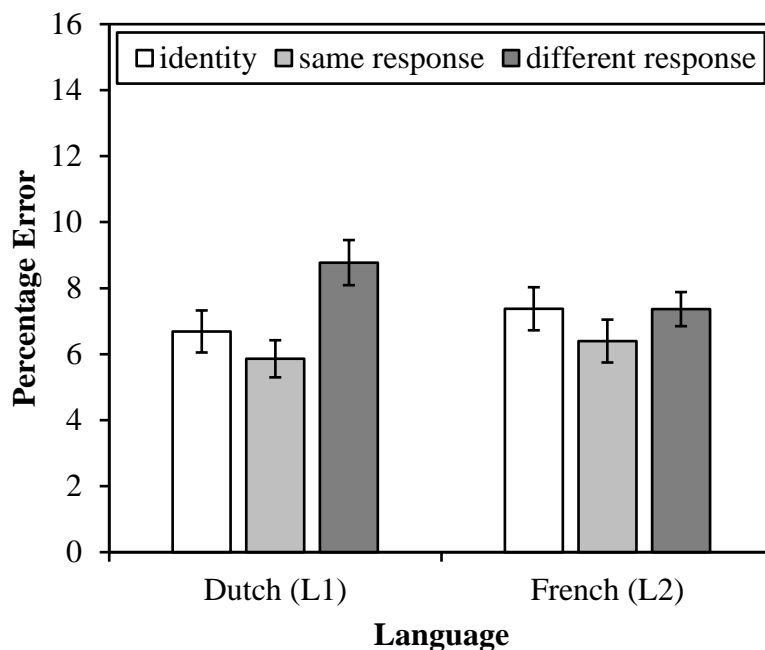

Figure 4. Percentage errors with standard errors for Dutch and French colour words.

Correlations

As a supplementary analysis, we consider how the measures of language level correlate with the stimulus and response conflict effects for both languages. Although the language measures correlate with each other well in the intuitive fashion (data available from the lead author on request), there was little evidence for a relationship between any of the language metrics with any of the observed congruency effects. The non-parametric Spearman's ρ correlations are presented in Table 2 (results were similar with the parametric Pearson's r). As can be observed, none of the performance (response time or error) measures correlated with years of French training, self-rated French level, or LEXTALE_FR score. With the LEAP-Q, percentage of Dutch and French language use also did not correlate with any performance measures. Age of (speaking) acquisition and fluency, and age of reading acquisition and fluency for both languages were not related to the response time or error effects after a Holm-Bonferroni correction for multiple comparisons. Without a correction, some correlations were significant at the $\alpha = .05$ level, but this should, of course, be interpreted with caution (e.g., the largest correlation is between age of beginning to learn to read French and the stimulus conflict effect in Dutch, which seems difficult to interpret). More generally, the relative lack of strong correlations between the conflict effects and the language skill metrics is probably not surprising given that there were little overall differences in the conflict effects across languages to begin with.

Table 2. Correlations with stimulus and response conflict effects.

	Dutch				French			
	Stimulus		Response		Stimulus		Response	
	RT	ERR	RT	ERR	RT	ERR	RT	ERR
LEXTALE_FR								
Years French	-.007	-.017	-.153	.099	-.029	-.163	-.137	.124
French Level	-.099	-.094	.022	.017	.018	-.116	-.117	-.007
Score	-.096	-.036	-.078	.161	-.081	-.176	-.175	-.138
LEAP-Q								
%Dutch Use	-.002	-.090	.041	.052	-.109	.060	-.165	-.021
%French Use	-.025	-.023	-.046	.002	.011	-.173	.062	-.060
Dutch								
Acquisition	-.059	.039	.085	*-.237*	.045	-.033	-.185	-.023
Fluent	.022	.065	.084	*-.283*	-.193	.140	-.005	-.039
Reading	.119	.061	.023	*-.265*	.063	.076	*-.240*	.039
Fluent Read	.146	.078	-.059	-.197	-.084	.079	-.083	-.007
French								
Acquisition	-.044	*.239*	.102	*-.271*	.024	.100	.079	-.097
Fluent	.032	*.210*	.070	-.110	-.054	.101	.130	.031
Reading	-.066	*.321*	.189	-.175	.036	.032	.035	-.002
Fluent Read	-.112	*.217*	.116	-.013	-.114	-.010	.176	.021

Notes: Italic = *p* < .05. No tests significant with Holm-Bonferroni correction.

Discussion

In the present report, we investigated for the first time the source of L2 (in addition to L1) congruency effects with a 2-to-1 keypress mapping procedure. Most importantly, the experiment revealed both stimulus and response conflict for a second language (French), just as with the first language (Dutch) in response times. This is contrary to the hypothesis that second language colour words act as mere associates for the first language translations, as colour associates (e.g., "sky") produce stimulus conflict alone (Schmidt & Cheesman, 2005). That is, the results are not consistent with the notion that second language colour words do not bias a potential response. Similarly, the current results are inconsistent with the notion that foreign language colour words only influence response selection (e.g., because they are not strongly enough connected to semantics). That is, the results are not consistent with the notion that second language colour words retrieve the response associated with the first language (e.g., via lexical translation), but do not activate semantics (or at least sufficiently to produce

stimulus conflict). Such a notion would also assume that retrieval of a response can bypass semantics (i.e., that "noir" can retrieve the black response without a mediation through semantics), which may or may not be plausible.

Also interesting, there were no sizeable differences in the observed congruency effects across languages, with the exception of some numerical trends and a marginally larger response conflict effect in errors for L1 colour words. As previously discussed, past reports have observed smaller interference effects from a second language than from a first language. Indeed, Mägiste (1984, 1985) argued that the amount of conflict is proportional to mastery of a language (see also, Brauer, 1998). This may seem inconsistent with the present report, if not for a few added considerations. First, the asymmetry between first and second language congruency effects partially depends on the response language (Atalay & Misirlisoy, 2012; Dyer, 1971; Preston & Lambert, 1969; Tzelgov et al., 1990). For instance, with verbal Dutch responses, Dutch colour word interference would increase and French colour word interference would decrease. The reverse would be true with verbal French responses. In the current experiment, keypress responses were used, which are not inherently compatible with either language. Thus, any observed asymmetry should not be expected to be particularly large. Future research directly comparing keypress and verbal response modalities might test this notion more directly (though unfortunately the 2-to-1 mapping procedure cannot be used with verbal naming responses).

There were some small hints of a larger congruency effect for L1 words in the current experiment, even marginally so in the error data (i.e., for response conflict). It might be supposed that Dutch-speaking participants sub-vocally name the colours in Dutch, making keypress responses more compatible with Dutch than French words. If this is true, however, it might also be surprising that an asymmetry with larger L1 conflict effects was not observed. On the other hand, the larger asymmetry in verbal experiments might have to do with the

much stronger stimulus-response compatibility for colour naming. Future research might investigate these possibilities more closely (e.g., by having participants sub-vocally name in Dutch versus French).

As one caveat with the present report, however, it is worth stressing that the present investigation made use of keypress responses, rather than verbal. This was necessary for the 2-to-1 mapping procedure. It remains possible, therefore, that additional differences between languages might be observable that are specific to the verbal response modality. For instance, during phonetic or articulatory planning, an L2 word like "jaune" might interfere less than an L1 word like "geel" because the former does not correspond to a potential response in the (Dutch language) response set. Indeed, this could additionally explain the dependence of the asymmetry in L1 and L2 Stroop effects on the response language discussed earlier. Future research may therefore be directed at disentangling these issues further.

For both the stimulus and response conflict effects, congruent trials were compared with incongruent trials. The conflict effects observed may therefore be in part due to incongruent-trial interference, and may in part be due to congruent-trial facilitation (Hasshim & Parris, 2014). Development of an appropriate neutral control condition (relative to which facilitation and interference can be measured), however, is a notoriously difficult task in Stroop and other research domains (Jonides & Mack, 1984; MacLeod, 1991). Future research may nevertheless aim to tease these subcomponents of the Stroop effect further apart in both L1 and L2 speakers.

Above caveats aside, the present results suggest that L2 colour words influence cognitive processing in much the same way as L1 colour words. L2 colour words produce both stimulus conflict and response conflict. Interestingly, this is even true for the non-cognates used in the present report. That is, foreign words (e.g., "marron") that *look* quite different from the native language equivalent (e.g., "bruin") automatically interfere with both

stimulus and response selection. Future research might aim to investigate the extent to which

the same is true for participants with considerably lower L2 language knowledge (e.g., no

direct formal training). For instance, in the Kroll and Stewart (1994) model discussed above,

it is assumed that the connections to semantics do increase with increasing proficiency.

Despite the case that French proficiency in our sample was relatively weak, it could be the

case that participants were proficient *enough* to induce stimulus conflict. On the other hand,

others have argued for relatively early semantic mediation (e.g., Duyck & De Houwer, 2008).

The approach to studying stimulus and response conflict in two languages novelly introduced

here might therefore be extended further to very early language learning to help in

discriminating between these competing ideas.

References

Altarriba, J., & Mathis, K. M. (1997). Conceptual and lexical development in second language acquisition. *Journal of Memory and Language, 36*, 550-568.

Atalay, N. B., & Misirlisoy, M. (2012). Can contingency learning alone account for item-specific control? Evidence from within- and between-language ISPC effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1578-1590.

Augustinova, M., & Ferrand, L. (2014). Social priming of dyslexia and reduction of the Stroop effect: What component of the Stroop effect is actually reduced? *Cognition, 130*, 442-454.

Augustinova, M., Silvert, L., Ferrand, L., Llorca, P. M., & Flaudias, V. (2015). Behavioral and electrophysiological investigation of semantic and response conflict in the Stroop task. *Psychonomic Bulletin & Review, 22*, 543-549.

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: Consequences for mind and brain. *Trends in Cognitive Sciences, 16*, 240-250.

Brauer, M. (1998). Stroop interference in bilinguals: The role of similarity between the two languages. In A. F. Healy & L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 317-337). Hillsdale, NJ: Erlbaum.

Brysbaert, M. (2013). LEXTALE_FR: a fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica, 53*, 23-37.

Chen, A. T., Bailey, K., Tiernan, B. N., & West, R. (2011). Neural correlates of stimulus and response interference in a 2-1 mapping stroop task. *International Journal of Psychophysiology, 80*, 129-138.

Chen, H. C., & Ho, C. (1986). Development of Stroop interference in Chinese-English

   bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition,*

   12, 397-401.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in

   the bilingual's two lexicons compete for selection? *Journal of Memory and Language,*

   41, 365-397.

Dalrymple-Alford, E. C. (1968). Interlingual interference in a color-naming task.

   Psychonomic Science, 10, 216-216.

De Houwer, J. (2003). On the role of stimulus-response and stimulus-stimulus compatibility

   in the Stroop effect. *Memory & Cognition, 31*, 353-359.

De Houwer, J. (2004). Spatial Simon effects with nonspatial responses. *Psychonomic Bulletin*

   & Review, 11, 49-53.

Duyck, W., & De Houwer, J. (2008). Semantic access in second-language visual word

   processing: Evidence from the semantic Simon paradigm. *Psychonomic Bulletin &*

   Review, 15, 961-966.

Dyer, F. N. (1971). Color-naming interference in monolinguals and bilinguals. *Journal of*

   Verbal Learning and Verbal Behavior, 10, 297-302.

Ferrand, L., & Augustinova, M. (2014). Differential effects of viewing positions on standard

   versus semantic Stroop interference. *Psychonomic Bulletin & Review, 21*, 425-431.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture

   processing. *Journal of Experimental Psychology: General, 118*, 13-42.

Hasshim, N., & Parris, B. A. (2014). Two-to-one color-response mapping and the presence of

   semantic conflict in the Stroop task. *Frontiers in Psychology, 5*, Article 1157.

Hasshim, N., & Parris, B. A. (2015). Assessing stimulus-stimulus (semantic) conflict in the

Stroop task using saccadic two-to-one color response mapping and prerespose

pupillary measures. *Attention Perception & Psychophysics, 77*, 2601-2610.

Jongen, E. M. M., & Jonkman, L. M. (2008). The developmental pattern of stimulus and

response interference in a color-object Stroop task: An ERP study. *BMC

Neuroscience, 9*, Article 82.

Jonides, J., & Mack, R. (1984). On the cost and benefit of cost and benefit. *Psychological

Bulletin, 96*, 29-44.

Klein, G. S. (1964). Semantic power measured through the interference of words with color-

naming. *American Journal of Psychology, 77*, 576-588.

Koda, K. (1996). L2 word recognition research: A critical review. *Modern Language Journal,

80*, 450-460.

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.

La Heij, W., Debruyn, E., Elens, E., Hartsuiker, R., Helaha, D., & Vanschelven, L. (1990).

Orthographic facilitation and categorical interference in a word-translation variant of

the Stroop task. *Canadian Journal of Psychology, 44*, 76-83.

Mackinnon, D. P., Geiselman, R. E., & Woodward, J. A. (1985). The effects of effort on

Stroop interference. *Acta Psychologica, 58*, 225-235.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative

review. *Psychological Bulletin, 109*, 163-203.

Mägiste, E. (1984). Stroop tasks and dichotic translation: The development of interference

patterns in bilinguals. *Journal of Experimental Psychology: Learning, Memory, and

Cognition, 10*, 304-315.

Mägiste, E. (1985). Development of intralingual and interlingual interference in bilinguals. *Journal of Psycholinguistic Research, 14*, 137-154.

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*, 940-967.

Posner, M. I., & Presti, D. E. (1987). Selective attention and cognitive control. *Trends in Neurosciences, 10*, 13-17.

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*, 295-301.

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*, 310-315.

Schmidt, J. R., & Cheesman, J. (2005). Dissociating stimulus-stimulus and response-response effects in the Stroop task. *Canadian Journal of Experimental Psychology, 59*, 132-138.

Schmidt, J. R., Cheesman, J., & Besner, D. (2013). You can't Stroop a lexical decision: Is semantic processing fundamentally facilitative? *Canadian Journal of Experimental Psychology, 67*, 130-139.

Schmidt, J. R., Crump, M. J. C., Cheesman, J., & Besner, D. (2007). Contingency learning without awareness: Evidence for implicit control. *Consciousness and Cognition, 16*, 421-435.

Schmidt, J. R., & De Houwer, J. (2012). Contingency learning with evaluative stimuli: Testing the generality of contingency learning in a performance paradigm. *Experimental Psychology, 59*, 175-182.

Simon, J. R., Craft, J. L., & Webster, J. B. (1973). Reactions toward stimulus source: Analysis of correct responses and errors over a five-day period. *Journal of Experimental Psychology, 101*, 175-178.

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: Effect of an irrelevant cue on information processing. *Journal of Applied Psychology, 51*, 300-304.

Smith, M. C., & Kirsner, K. (1982). Language and orthography as irrelevant features in color-word and picture-word Stroop interference. *Quarterly Journal of Experimental Psychology, 34A*, 153-170.

Stirling, N. (1979). Stroop interference: An input and an output phenomenon. *Quarterly Journal of Experimental Psychology, 31*, 121-132.

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

Tzelgov, J., Henik, A., & Leiser, D. (1990). Controlling Stroop interference: Evidence from a bilingual task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 760-771.

van Veen, V., & Carter, C. S. (2005). Separating semantic conflict and response conflict in the Stroop task: A functional MRI study. *Neuroimage, 27*, 497-504.

Werker, J. F., & Byers-Heinlein, K. (2008). Bilingualism in infancy: First steps in perception and comprehension. *Trends in Cognitive Sciences, 12*, 144-151.

Appendix: Language Demographics

There was relatively little variability in the language demographics of the participants. The vast majority of participants rated their order of language acquisition as Dutch, followed by French (followed by English), but rated their language dominance as Dutch, followed by English, followed by French in third, and indicated Dutch as their native language. Most critically, all but two participants ranked Dutch as their dominant language. One of the remaining two indicated that both Dutch and Turkish were joint dominant languages, and the other indicated Russian as the dominant language. Both, however, ranked Dutch as second and French as fourth. One other participant rated both Dutch and French as equally dominant (though with Spanish and English as native but not dominant languages), but the results on the LEXTALE_FR did not support this. Though this participant had the highest score in the sample (tied with another) of 32, this only corresponds to the 4th percentile for native French speakers. Additionally, this participant indicated that they used Dutch much more frequently (75%) than French (5%). Another two participants indicated French as a joint first language (one with Dutch and the other with Turkish) who were also among the higher scorers on the LEXTALE_FR (27 and 32, respectively). However, none of these participants rated French as their dominant language and, as already mentioned, none of the participants in the sample had particularly convincing LEXTALE_FR scores. On average, participants self rated their French fluency at 6.3 on a 0-10 scale. In contrast, the average LEXTALE_FR score was only 5.8 (out of a maximum score of 56), corresponding to the 1st percentile for L1 users and 48th percentile for L2 users. All participants gave a higher percentage for Dutch language use than for French, with one exception. One participant rated Dutch and French both 50%, though, strangely, this participant rated French as their third most dominant language, self rated their French knowledge as 6 (average), and only scored 1 on the LEXTALE_FR. All in all, the participants in the sample seemed to fit the basic requirements of dominant Dutch and less

dominant French. Thus, no participants were trimmed. Trimming of the potentially

problematic participants mentioned above had only negligible impact on the results reported

below.

The vast majority of participants correctly translated "noir(e)" (90/93), "vert(e)"

(88/93), and "jaune" (83/93). For "brown," most participants indicated "brun(e)" (66/93), with

only a few indicating "marron" (7/93, two of which wrote both "marron" and "brun"). "Brun"

is more similar to the Dutch "bruin" (and English "brown"), but is only a partially correct

translation for the colour brown.[3] Any incorrect answers (including misspellings) were

pointed out to the participant before starting the experiment, including the semi-

mistranslations of "brown." Note that though very few participants indicated "marron," most

participants did seem to recognize this word as soon as it was presented to them by the

experimenter (this was not tested systematically, however).

---

[3] "Brun" is primarily srused for hair colours, and derivatives (e.g., "ours brun" [brown bear], which has brown hair, or "bière brune" [brown beer], which also relates to hair in the same was as "bière blonde" [blonde beer]). "Brun" can also be used to refer to specific shades or in an "artistic" context (e.g., like referring to blue as "azure" in English). "Marron" is the more standard colour name for brown.