THESE DE DOCTORAT DE L'ETABLISSEMENT UNIVERSITE BOURGOGNE FRANCHE-COMTE

Ecole Doctorale n°554

Ecole doctorale Environnements – Santé

Doctorat de Psychologie Cognitive

Par


Iva Šaban


# Basic processes in interference paradigms

Thèse présentée et soutenue à Dijon, le 2 décembre 2021

Composition du jury :

| | | |
|---|---|---|
| Prof. AUGUSTINOVA Maria | Université de Rouen Normandie | Rapporteur |
| Prof. FERRAND Ludovic | Université Clermont Auvergne | Rapporteur |
| Prof. DE HOUWER Jan | Ghent University | Examinateur |
| Prof. LEMERCIER Céline | Université Toulouse Jean Jaurès | Examinatrice |
| Prof. POULIN-CHARRONNAT Bénédicte | Université de Bourgogne | Examinatrice |
| Dr. SERVANT Mathieu | Université de Franche-Comté | Examinateur |
| Prof. SCHMIDT James R. | Université Bourgogne Franche-Comté | Directeur de thèse |

ÉCOLE DOCTORALE
PRES I Bourgogne I Franche-Comté
Environnements - Santé

**Titre :** Processus basiques dans les paradigmes d'interférence

**Mots clés :** bilangue, l'effet Stroop, conflit stimulus, conflit réponse, l'apprentissage des mots, l'apprentissage d'une langue étrangère

**Résumé :** Dans la tâche Stroop, les sujets prennent plus de temps pour nommer la couleur d'un mot incongruent (« rouge » écrit en bleu), que d'un mot congruent (« rouge » écrit en rouge). Ce phénomène est connu comme l'effet Stroop. Il est observé avec les mots de la langue maternelle ainsi qu'avec les mots d'une langue étrangère. L'objectif de cette thèse est d'étudier l'origine de cet effet dans le cas d'une langue moins maitrisée.

Le premier composant de cet effet, appelé conflit stimulus est produit par la différence entre la signification du mot et sa couleur. Le deuxième composant, appelé conflit réponse provient du conflit entre les deux réponses possibles. La littérature suggère que ces deux conflits contribuent à l'effet Stroop et peuvent être étudiés séparément par la procédure de *dissociation de la touche 2-en-1*.

Selon des théories du développement linguistique, un seul conflit devrait se produire pour une langue étrangère peu maitrisée. Cependant, la première expérience a montré la présence de ces deux types de conflit dans cette situation. Une série d'expériences a ensuite cherché à tester la présence de ces conflits dans le cas de mots issus d'une langue inconnue. Les sujets ont été entrainés avec des mots croates associés avec leur traduction et leur représentation sémantique. Les entraînements variaient dans leur structure (e.g., le type d'essai, le nombre de réponses alternatives, etc.), leur durée (de 32 à 576 essais) et le type de mots entrainés (mot désignant une couleur ou mot associé à une couleur). Suite à ces entraînements, les participants ont effectué la tâche Stroop.

Nos résultats montrent un conflit stimulus au niveau du temps de réaction et un conflit réponse au niveau des erreurs dans le cas de la langue récemment apprise avec un entraînement optimal (Expérience 4, comparé aux Expériences 2 et 3 avec une phase d'entraînement plus courte). En revanche, aucun conflit n'a été observé avec les mots associés à des couleurs (Expérience 5 et 6). Cela signifie que lorsqu'ils sont suffisamment bien appris, les mots d'une langue inconnue peuvent influencer l'identification sémantique (conflit stimulus) et la sélection de réponse (conflit réponse).

ÉCOLE DOCTORALE
PRES I Bourgogne I Franche-Comté
Environnements - Santé

**Title :** Basic processes in interference paradigms

**Keywords :** bilingualism, Stroop effect, stimulus conflict, response conflict, novel word acquisition, foreign language learning

**Abstract :** The aim of the present thesis is to investigate the source of Stroop (interference) effects in weak bilinguals (Experiment 1) and in early language learning (Experiment 2-6). Participants performed a bilingual colour-word Stroop task with intermixed first language (L1) and second language (L2) words. The typical finding from the Stroop literature is slower and less accurate responding when the word and colour are incongruent (e.g., “red” in blue) relative to congruent (e.g., “red” in red). Interestingly, this congruency effect occurs for the colour words from both L1 and L2. What produces this congruency effect? That is, what is the source of the conflict produced by incongruent colour words? First, stimulus or semantic conflict is a conflict between the meaning of the word and ink colour. Second, response conflict occurs when different response alternatives are activated. Both types of conflict contribute to L1 congruency effects.

According to some theoretical accounts on early language learning, only one of these two types of conflict should emerge for non-fluent L2. Stimulus and response conflict are separated with a 2-to-1 keypress dissociation procedure. Both stimulus and response conflict were evidenced for the weakly spoken L2 (Experiment 1; English in native French speakers). In series of L2 word learning studies, participants were trained with novel Croatian colour words associated with their L1 translations and corresponding semantic representations. Word trainings differed in their structure (types of training trials, number of response alternatives), length (from 32 to 576 trials) and to-be-learned word types (colour words, colour associates) across studies. The L2 word trainings were followed by the Stroop task. Stimulus conflict was observed in response times and response conflict in errors for recently learned L2 words (Experiment 4) when optimal training was administered (in contrast to Experiment 2 and 3, with considerably shorter training). However, this approach did not reveal the source of conflict with colour associates, because no substantial L2 Stroop effect was observed for these stimuli (Experiment 5 and 6). The present findings suggest that low proficient L2 words, when trained in adequate conditions, are potent enough to affect semantic identification and response selection.

*Mojim roditeljima, Arijani i Ivici Šaban*

**Acknowledgement**

This research is supported by the French “Investissements d’Avenir” program, project ISITE-BFC. I thank my supervisor, James R. Schmidt, for his thoughtful comments, constructive feedback and kind support during my PhD.

A kind thank you to my committee members: Jan de Houwer, Céline Lemercier, Bénédicte Poulin-Charronnat and Mathieu Servant for agreeing to participate in the defense of this thesis. I would like to thank Maria Augustinova and Ludovic Ferrand, for their very helpful comments and suggestions.

Very special thanks to the members of the LEAD for creating a pleasant working environment. Last but not least, thanks to my colleagues and friends from the lab, for a cherished time spent together in the lab and in social settings.

# Contents

# List of figures

# List of tables

## Introduction

### 1. Stroop task

The Stroop task is the commonly adopted name for the serial colour-word task reported by J. R. Stroop in 1935. The original task required participants to read through lists of colour words (e.g., “blue”, “green”, etc.) printed either in neutral black ink or coloured ink on cards. Results of the first experiment were not surprising; participants were equally fast in reading the colour words printed in black as they were when reading the words coloured in the ink different from the meaning of the word. This indicated that the print colours do not influence reading performance, likely because we read words much faster than we can name colours (Cattell, 1886; Fraisse, 1969). The second experiment, however, yielded interesting results; Stroop demonstrated that participants are slower in naming the colour of the ink of colour words that are incompatible than in naming the colours of squares or other colour-irrelevant symbols.

Most modern Stroop experiments use a computer instantiation where response times and errors are measured for each individual stimulus. In these studies, the *congruency* or *Stroop effect* is typically defined by the finding that participants respond faster and more accurately when the colour word and ink colour are *congruent*, where the word and colour match in meaning (e.g., the word “red” printed in red), relative to *incongruent*, where the word and colour mismatch (e.g., the word “red” printed in blue; for a review, see MacLeod, 1991). The Stroop effect had soon become the hallmark of interference paradigms, inspiring a huge number of researchers worldwide and shaping theoretical accounts over the years.

Although participants are explicitly instructed to focus on colour identification as the relevant task and ignore the distracting word, they are not completely able to fulfil these requirements. To measure the effects of distracters on responses in the Stroop task, researchers typically pool the response latencies on congruent (e.g., “red” in red), incongruent (e.g., “red” in blue), and neutral stimuli (e.g., “xxxx” in red) separately. By subtracting the average response latency to neutral stimuli from the latencies to the incongruent and congruent stimuli, they measure for inhibition and facilitation, respectively (Glaser & Glaser, 1989). For instance, in the incongruent condition (e.g., “red” in blue), the conflict between the irrelevant meaning of the word and ink colour needs to be resolved. Therefore, the distracter has an inhibitory (or interfering) effect, and it delays the response. Cognitive control and conflict resolution processes need to be engaged, which results in delayed response times compared to the neutral condition (e.g., “xxxx” in red) where this conflict does not occur (Stroop, 1935). Stroop facilitation refers to the faster response times in the congruent condition (e.g., “red” in red) than in the control condition (e.g., “xxxx” in red). In this example, the congruent distracter refers to the same colour as target (i.e., ink colour), which facilitates colour identification (Glaser & Glaser, 1989). These two phenomena, inhibition (or interference) and facilitation, have been discussed in the Stroop literature in terms of their asymmetry. In particular, facilitation is typically much smaller than interference (MacLeod, 1998).

Apart from the relative magnitude of interference and facilitation in the Stroop task, these effects have often served as evidence for the automaticity of reading and its interaction with other cognitive processes. That is, the tendency to read words when they are presented to us is so automatic that we cannot help but (at least partially) read the word, even when instructed not to. Much debate has focused on the automaticity of word reading, with a body of evidence suggesting that word reading is automatic and cannot be prevented (see Augustinova & Ferrand, 2014). In contrast, some other findings indicated that word reading is controllable and that the automaticity of reading is only a myth (see Besner et al., 1997). Apart from this debate, a substantial body of literature (MacLeod, 1991) has recognized

Stroop interference as a complex phenomenon influenced by many cognitive mechanisms, thus generating the questions about its origin. For instance, much debate has focused on investigating which processes could possibly explain this interference. In the present thesis, I focus on semantic and response processing (De Houwer, 2003) and their separate contributions to overall Stroop interference, discussed in the following section.

***Conflicts in the Stroop task***

As already mentioned, one question of interest in the Stroop literature is the source of the conflict. What produces the conflict between a word meaning and an incongruent ink colour? In this section, I briefly discuss a historical overview concerning the conflicts in the Stroop task.

First, conflict in the Stroop task could occur between the meaning of the word and its colour, termed as *stimulus* or *semantic conflict* (Glaser & Glaser, 1989; MacKinnon et al., 1985). This conflict is assumed to occur early, right after perceptual encoding, when the to-be-processed information gets in touch with semantic memory. This was supported by the *semantic-encoding hypothesis* of Seymour (1974, 1977). Seymour (1974) presented subjects with a word (distracter) placed either below or above a square. Four pairs of distracters were used: “yes”/”no”, “right”/”wrong”, “up”/”down”, and the control or neutral stimulus “xxx”. An instruction word printed inside the square was either “above” or “below”. The subjects responded “true” when the distracter corresponded to place indicated by the instruction word, and “false” when it did not. In a control condition, above displays were classified faster than below displays and true displays faster than false displays. When the distracters “yes” and “no”, and “right” and “wrong” were used, Seymour expected them to facilitate and interfere with the responses “true” and “false”. Surprisingly, no interference effect was found. However, when distracters were “up” and “down”, a substantial Stroop effect was observed. That is, interference emerged only when the meaning of distracter (i.e., “up” or “down”) differed from the meaning of the instruction adverb (i.e., "above" or "below"). There was a competition between two semantically related words that are used as input in the word-picture comparison task (Seymour, 1974). In another experiment, more similar to the standard Stroop task, Seymour (1977) revealed that the conflict is located at the level of the semantic representations of the relevant and the irrelevant stimulus dimensions. The simultaneous activation of two closely related semantic dimensions leads to a conflict that must be resolved before further processing (Seymour, 1977). In other words, semantic or stimulus conflict (SC) appears due to the crosstalk between relevant (i.e., ink colour) and irrelevant dimensions (i.e., word meaning) in the Stroop paradigm.

Glaser and Glaser (1989) introduced a model of the Stroop interference effect that assumes the existence of two separate systems: a *semantic system* that contains semantic knowledge, and a *lexical system* that contains linguistic knowledge. According to their model, the semantic system controls the perception of visual stimuli (e.g., colours, pictures) and responding to them (e.g., key responding), while the lexical system allows vocal and written production (e.g., colour naming). Therefore, colours and pictures share privileged access to semantic information and words have privileged access to the mental lexicon. This model assumes the occurrence of Stroop interference if the distracter has direct access to the system that is critical for task execution. For instance, when the ink colour has to be named (i.e., the standard Stroop task), the colour activates corresponding nodes in the semantic system, followed by the word nodes in the lexical system. For instance, presenting the word "red" in blue automatically activates the concept "red" that is retrieved from lexical system. This interferes with the activation of the concept "blue" that corresponds to the relevant task dimension. Therefore, word meaning and ink colour meaning in incongruent trials compete for processing resources (e.g., in semantic memory), which impairs colour identification (Glaser & Glaser, 1989).

Second, another account of the hypothesized source of Stroop effect concerns response processing. The *response competition hypothesis* (Klein, 1964; Morton, 1969; Posner & Snyder, 1975) argues that words and colours are processed in a parallel manner before activation of their motor programs. In the standard colour-word Stroop design, responses to each colour are assigned to different keys. In incongruent trials (e.g., "red" in blue), both dimensions of Stroop stimuli activate a potential response, one corresponding to the ink colour and another corresponding to the word. Since only the motor program that corresponds to relevant stimulus dimension ("blue") is admitted to execution, the program that corresponds to the irrelevant dimension ("red") must be prevented. This produces interference at the level of response output (De Houwer, 2003). In other words, the response evoked by the word (e.g., red) and the response evoked by the colour (e.g., blue) compete for selection, which produces *response conflict* (Klein, 1964; Morton, 1969; Posner & Snyder, 1975; Posner & Presti, 1987).

For instance, Klein (1964) was the first to investigate whether written distracters unrelated to colours influence the magnitude of the Stroop interference. Klein demonstrated that the Stroop interference does not occur only when the distracting colour word and the ink colour are incompatible, but it can also occur with other written stimuli as distracters. These written stimuli that represent distracters in the Stroop task differ in the degree of semantic overlap with the colour word (target). This has been described as evidence for a *semantic gradient* in the Stroop effect. That is, an increase in the semantic relationship between the distracting word and print colour produces a larger Stroop interference effect (Klein, 1964).

In his experiments, a set of four colour-words was used (red, blue, green, and yellow). The distracter set consisted of either nonsense-syllables ("evgjc", "bhdr"), rare words ("sol", "helot"), common words ("friend", "heart"), colour-associated words ("lemon", "sky", which are related in meaning to yellow and blue, respectively), colour words that do not belong to the response set (“purple”, “grey”, which are not potential target colours), and the colour words from the response set (“red”, “blue”). Colour naming speed was measured in these six conditions. The nonsense syllables produce the lowest magnitude of interference. The largest interference was found when the distracter came from the same response set (e.g., “red” in blue), that is, with typical Stroop stimuli. If the distracter word is a colour word but one that does not originate from the response set (“purple”), interference was dramatically reduced. Even less interference is produced when the colour associates were used as distracters (“lemon”, printed in blue). When the distracter was a common, colour-unrelated word (“friend”), the interference was minimal (Klein, 1964). According to Klein, these results support a notion that different word types influence response selection as a function of the *semantic gradient*. That is, the more a distracting word is related in meaning to a target colour, the more interference is produced.

Sharma and McKenna (1998) argued that the multi-component nature of the semantic gradient could be explained by the series of separate interference effects. They used five types of stimuli: letter strings (“xxxx”, “hhhh”, “oooo”), neutral words (“top”, “club”, “stage”), colour related words (or “colour associates”; “fire”, “grass”, “sky”, related in meaning to red, green, and blue, respectively), colour words out of the response set (“purple”, “grey”, “yellow”), and incongruent colour words (“red”, “blue”, “green”). They argued that these five conditions allowed them to analyse the separate contribution of four components of Stroop interference. For instance, a lexical component was computed by subtracting the response time on letter strings from response time on neutral words. The semantic relatedness was computed by subtracting neutral words from colour-related words. The semantic relevance is a result of subtracting colour-related associates from colour words out of the response set. Finally, the response set membership effect was accessed by subtracting the nonresponse colour word types from the incongruent trials (i.e., colour words in the response set). These four components were examined for manual (i.e., keypress) and vocal (i.e., colour naming) response modalities, which varied as a within-subject factor: participants were instructed to ignore the distracter and to either name the colour aloud (vocal response) in one block or identify it by keypress (manual response) in another block. The overall interference effect (incongruent words – letter strings) was significant for both response modalities, with a larger interference effect produced for vocal responses. In the vocal variant of the Stroop task, all four components were present. In contrast, in the manual variant of the task, only response set membership produced a significant interference effect, with no significant effects for the lexical component, semantic relatedness, or semantic relevance. Response set membership was the only evaluated component that produced a larger interference effect for manual than for vocal responses (Klein, 1964). Sharma and McKenna concluded that the lexical/semantic component does not produce Stroop interference effect when manual responses are required. In other words, they claimed that semantics play role in colour identification but only for vocal responses. However, this final conclusion has not gone without critique: A reanalysis of their data showed that the semantic component of the interference effect occurs also in the manual version of the Stroop task (Brown & Besner, 2001).

To sum up, various authors have discussed the relative contribution of stimulus and response conflict effects for different word types (i.e., colour words, colour associates, etc.) in explaining the source of the interference effect. Nowadays, stimulus (semantic) and response conflict effects are the most dominant theoretical accounts of the Stroop interference (for others, see *General Discussion*). After years of debate, the current consensus is that both stimulus and response conflict effect contribute to the standard Stroop effect (Augustinova & Ferrand, 2014; Ferrand & Augustinova, 2014). There are several lines of evidence that converge on this notion, one of which is discussed in the following section.

### *2-to-1 mapping procedure and three types of trials*

A clear evidence for the contribution of both stimulus and response conflict effects comes from 2-to-1 mapping experiments. De Houwer (2003) introduced a new variant of the Stroop task with manual responses that allows the researcher to separate semantic and response conflict. In this paradigm, four colour words and their corresponding colour names (e.g., “blue”, “red”, “yellow”, “green”) are used. This manipulation assigns two possible response colours to one key (e.g., “blue” and “green”), and another two response colours to another key (“red” and “yellow”), as illustrated in Figure 1.

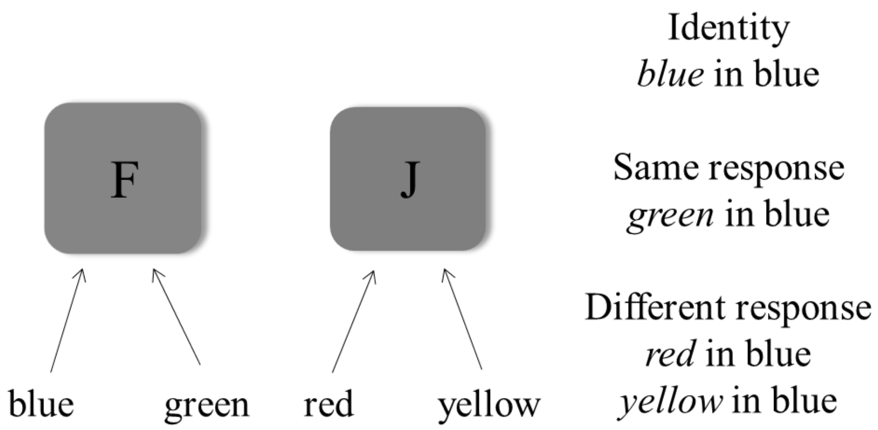


***Figure 1.*** Illustration of the 2-to-1 mapping procedure

Using this manipulation, three types of trials can be identified. First, there are *identity* trials (e.g., “blue” in blue), which are effectively the typical congruent trials in which the word matches ink colour. They are both semantically and response compatible since the meaning of the word corresponds the ink colour and the responses associated with both the word and colour are assigned to the same key. Second are *same response* trials (e.g., “green” in blue) which are incongruent in meaning (i.e., stimulus-incompatible; green and blue are two different colours) but are mapped to the same key (i.e., response-compatible; green and blue are assigned to the same key, “F” in Figure 1). A difference in response latencies between identity and same response trials corresponds to the stimulus conflict effect, as the two differ in stimulus compatibility (compatible and incompatible, respectively), but do not differ in response compatibility (i.e., because both are response compatible). Third, there are *different response* trials (e.g., “red” in blue) in which the word mismatches the ink colour (i.e., stimulus-incompatible) and the word and the colour are assigned to different response keys (i.e., response-incompatible). For instance, red corresponds not only to a different colour concept than blue, but red and blue are also assigned to different response keys (“J” and “F”, respectively, in Figure 1). A difference between same response and different response trials therefore indicates a response conflict effect. Typically, same response trials are slower than identity trials (indicating stimulus conflict), but faster than different response trials (indicating response conflict).

***Colour associates***

One evidence for the effect of semantics on colour identification in Stroop-like tasks with manual responses comes from colour associates. As briefly mentioned above, colour associates (e.g., “fire”) are words that are closely related to colour words (e.g., “red”) and their underlying semantic representations (Tanaka & Presnell, 1999). Colour associates have been found to produce interference with colour naming in the Stroop task. Like colour words, colour associates can be either congruent (e.g., “fire” printed in red) or incongruent (e.g., “fire” printed in blue) with the associated ink colour. By contrasting the two, a Stroop-like congruency effect occurs, with slower and less accurate responding to incongruent colour associates relative to congruent colour associates (Glaser & Glaser, 1989; Klein, 1964; Risko et al., 2006; Schmidt & Cheesman, 2005). There are three separated accounts that have discussed the source of interference effect in colour associates.

First, the difference in performance observed on congruent versus incongruent colour associates has been argued to be due to early, semantic processes rather than late, response processes (Glaser & Glaser, 1989). The reasoning for this explanation is based on the association between the two stimulus dimensions. When the distracter word (e.g., “fire”, associated to red) is printed in an incongruent colour (e.g., “blue”), two competing semantic colour representations (i.e., red and blue) are simultaneously activated, producing stimulus conflict. In other words, colour associates slow down the identification of an incongruent ink colour even when the meaning of the distracter word was assigned to the same key. On the response level, there is no direct relationship between the colour associates and colours (e.g., “fire” is not a potential response). Therefore, the response sets evoked by the target colour (“blue”) and the distracter (“fire”) are different, producing no response conflict. According to this (perhaps more philosophical) perspective, colour associate Stroop effects support the notion that, at least to a certain degree, the Stroop effect results from early, semantic processes.

Second, not all researchers agree with the semantic account of the colour associate effect. Klein (1964) claimed that colour associates might have a direct effect on response output by producing the colour response assigned to the colour associate. For instance, when the word “fire” is printed in blue, both the response linked to the colour red (i.e., the colour associated with “fire”) and the response linked to the colour blue (i.e., which is associated to the target colour) will be activated. This will, according to this perspective, produce response competition, resulting in response conflict exclusively, rather than semantic conflict (Klein, 1964).

Third, Sharma and McKenna (1998) suggested that colour associates effects are due to interference at the lexical level and not at the semantic level. This account was supported by the fact that they observed an interference effect for colour associates that only occurs when vocal responding is required, as previously discussed. The effect was eliminated when manual responses were used instead of vocal responses. Thus, according to those authors, semantic and response conflict effects are not potential sources of the interference effect in colour associates. Sharma and McKenna (1998) concluded that the effect of colour associates on

Stroop task performance is limited to lexical processing. However, these effects will not emerge using manual responses, since the vocal system does not control motor (key press) responses.

### ***Conclusions***

To sum up, the source of Stroop interference has been in focus of researchers over many decades, aiming to clarify the role of semantic and response processing. However, current findings mostly concern colour identification in native language (i.e., first language; L1) words, with relatively little attention on foreign language (i.e., second language; L2) words. To provide a broader framework regarding word processing in L2, in the next chapter I discuss the theoretical accounts of bilingual memory organization and related empirical findings.

## 2. Bilingualism

Over the past several decades, bilingual cognition has attracted increasing attention from researchers (Bialystok et al., 2012; Grosjean, 1997, 2012). One critical issue in the bilingual cognition literature concerns the interaction between the languages, that is, how the activation of one language system influences functioning of another language system. This interaction reflects on cognitive processes in bilinguals, and leads to another question of interest: whether there are similarities in the way the first language (L1) and second language (L2) influence cognitive processing?

The representation and structure of languages in bilinguals are a good starting point in discussing these issues. For instance, a basic feature of being bilingual concerns possessing multiple lexical representations (i.e., one for each language) for a particular concept (e.g., *green* and *vert* are the English and French words, respectively, for the same colour concept). However, one might ask whether and how these lexical representations are connected between themselves and with their corresponding semantic concepts. One account suggests that lexical representations from different languages are stored in separate mental lexicons (i.e., a separate mental "dictionary" for each language) and corresponding concepts are connected at a semantic level. A lack of repetition priming effects across languages supported this account (Kirsner et al., 1980, 1984; Scarborough et al., 1984). In one of these experiments, a group of Hindi-English bilinguals performed a lexical decision task on a set of 22 words and 11 non-words in either Hindi or English. In the following phase, the original words were repeated in either the same or in the alternative language, together with 22 new words and 22 non-words. They observed a facilitation in response latencies when words were repeated in the same language (i.e., English-English and Hindi-Hindi), with little or no facilitation in the between-language condition (i.e., English-Hindi and Hindi-English). They concluded that word representations at the lexical level are stored in separate structures (i.e., they are language-specific) in bilinguals (Kirsner et al., 1980). In contrast, another account assumes that words from all languages are stored in a shared structure (Paivio et al., 1988). The evidence for this account comes from semantic priming studies (Keatley et al., 1994; Tzelgov & Eben-Ezra, 1992). The processing of a target word (e.g., “nurse”) is faster and more accurate when it is preceded by semantically related prime word (e.g., “doctor”) than when it is preceded by unrelated prime (e.g., “table”; Meyer & Schvaneveldt, 1971). Semantic priming has also been investigated both within (i.e., prime and target are from the same language) and between languages (i.e., prime and target are from different languages). Numerous studies evidenced equal within- and between-language priming effect regardless of stimulus onset asynchronies (Tzelgov & Eben-Ezra, 1992) or difference in the orthographies between two languages (Chen & Ng, 1989; Tzelgov & Eben-Ezra, 1992). These results support the notion that the semantic store is not organized by languages, but rather shared by different languages. The existence of a common semantic (conceptual) store shared by different languages raises an issue of lexical access. That is, how do we access words from different languages stored in our lexicon? For instance, when presented with a certain word, do we first access the lexicon from one language and then from the second one, or do we parallelly search through all languages (de Bot, 2004)?

To reconcile different accounts and empirical findings, researchers have suggested that the bilingual memory system consists of both a lexical level in which lexical representations (i.e., words) are stored, and a conceptual level (i.e., semantics) in which the semantic representations of the words are stored. The nature of the connections *between* these lexical and conceptual memory stores was still unclear and at the centre of much debate. Different authors suggested their own view of the interaction between different memory stores, which resulted in the creation of several models of the bilingual memory system (Chen & Leung, 1989; de Groot, 1992; Kroll & Stewart, 1994; Potter et al., 1984).

***Organization of bilingual memory***

Potter and colleagues (1984) contrasted two models that aimed to explain how people store and process words in their second language. These models are based on the common assumption of two levels of representation which are hierarchically organized (Potter et al., 1984; Snodgrass, 1984). Words in a bilingual's memory are stored in separate lexical stores at the lexical level (i.e., one for each language). Concepts are stored in a common (i.e., shared) abstract memory store at the conceptual or semantic level. According to the two models of bilingual memory (see below), the levels of bilingual representation (lexical and semantic) are interconnected differently.

First, the *word association* model suggested the existence of direct associative links between words in two languages. For instance, a native English speaker that studies French will learn a foreign word "vert" in association with its native language equivalent, "green". A direct association is established between the first language (L1) and the second language (L2) lexicon. However, only the L1 words are associated with the corresponding concepts directly (Potter et al., 1984), as presented at Figure 2. For instance, "vert" which is learned as a translation of "green" and can only arrive at the semantic representation of "green" via the indirect connection through the first language equivalent. Empirical findings regarding this account will be explained shortly.

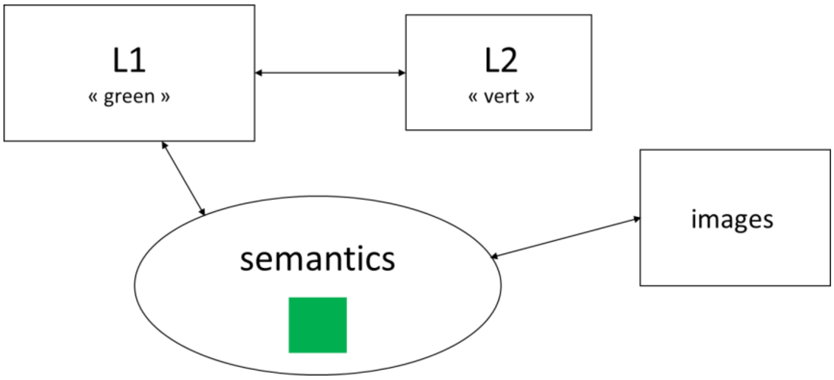


***Figure 2.*** Illustration of the Word Association model

Second, the *concept mediation* model, assumes that first and second languages operate independently at the lexical level. In other words, the words from two languages are not directly associated, but two lexicons share a common non-linguistic conceptual system. For instance, a native English speaker links a novel word "vert" to the conceptual representation of the colour green, rather than to its first language translation equivalent, as visible at Figure 3. Examples of the research that tested these accounts is provided in the subsequent paragraph.

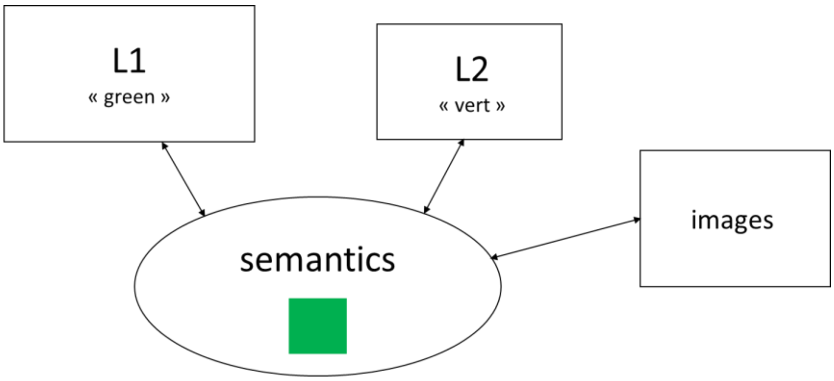


***Figure 3.*** Illustration of the Concept Mediation model

To test these hypotheses, groups of fluent Chinese-English bilinguals and non-fluent English-French bilinguals performed three tasks: picture-naming in both languages, word-reading in the first language (less relevant for the present discussion), and word-translating from first to second language (L1-L2). Two models provide different predictions about how subjects respond to words and pictures when L2 is a response language. The word association model assumes that L1-L2 word translating requires three steps: recognition of the L1 word, retrieval of the corresponding L2 word, and pronouncing the L2 word. On the other hand, picture naming in L2 includes five steps: picture recognition, concept retrieval, retrieval of the L1 word, retrieval of the L2 word, and pronouncing the L2 word. Considering the processing times for different mental steps in L1-L2 translating and L2 picture naming, the word association hypothesis assumes that picture naming in L2 is slower that L1-L2 word translating.

On the other hand, the concept mediation model predicts that these two tasks require similar mental steps. For instance, the translating task should involve recognition of the L1 word, concept retrieval, retrieval of the L2 word, and pronouncing the L2 word. In the picture-naming task, the following processing steps are adopted: picture recognition, concept retrieval, retrieval of the L2 word, and pronouncing the L2 word. Thus, according to this account, word translating and picture naming require an equal degree of semantic access, which results in approximately the same time required to complete both tasks. The results of Potter and colleagues (1984) showed that when the response language was L2, no difference in the response speed between picture naming and L1-L2 translating was observed. This pattern was evidenced for both groups of subjects. According to the authors, there is no direct association between words in the two languages, with L2 words being mapped to corresponding L1 words via a common conceptual system, even for non-fluent bilinguals (Potter et al., 1984).

One critique of these results concerns the level of second-language proficiency within the sample. For instance, Chinese-English bilinguals received different amount of L2 training, while English-French bilinguals, even though not fluent in their L2, possibly had sufficient L2 knowledge to produce these effects. Chen and Leung (1989) tested these findings in subjects with different degrees of L2 proficiency (i.e., proficient bilinguals, newly adult beginners, second- and fourth-grade beginners) who performed the word reading, picture naming, and word translating tasks. When the response language was L2, the proficient bilinguals were equally fast in both translating and picture naming tasks. However, adult L2 beginners were more efficient in translating than picture naming. For child beginners, picture naming was faster than translating. These results are consistent with an idea that beginners at early stages of L2 acquisition rely on L1 translation equivalents (i.e., the adult learners) or pictorial representations (i.e., the child beginners), whereas proficient bilinguals tend to directly access the meanings of L2 words. The L2 proficiency, as well as the age or method of L2 acquisition, play important roles in the determining the pattern of L2 lexical processing (Chen & Leung, 1989).

***The Kroll and Stewart (1994) model***

Since the two above-described models tap into bilingual populations of different L2 proficiency, Kroll and Stewart (1994) merged the two into one larger theory, known as the *Revised Hierarchical Model (RHM)*. It has become one of the most important and frequently cited models of bilingual memory representation. The Kroll and Stewart (1994) model is in its nature a developmental model focused on the interlanguage connections between the lexical and conceptual representations in the early stages of L2 acquisition (Kroll & Stewart, 1994).

The most evident difference from the prior models is the assumption that the bilingual's lexicons are bidirectionally linked, as shown at Figure 4. The lexical links between two lexicons have different strengths: the link from the L2 to the L1 lexicon is stronger than vice versa. This reflects the fact that during second language acquisition, bilinguals learn L2 words as direct translations of familiar L1 words (e.g., that "vert" is the French word for "green"). This results in the formation of a lexical-level association that remains strong and active in one direction (L2-L1), but weaker in the other (L1-L2). The model further proposes bidirectional conceptual links between the two lexicons and the conceptual store. The conceptual store is an integrated, common store for both languages (e.g., one conceptual representation for the “green” concept, which is linked to both the “green” and “vert” lexical representations). There is a strong conceptual link between the L1 lexicon and the conceptual store, explained by the privileged status of L1 in accessing the concepts. The conceptual link between the L2 lexicon and the conceptual store is relatively weak, representing the bilingual’s inability to access the concepts directly from the L2 lexicon. Considering the difference in strength of the conceptual links between the lexicons and semantics, the activation of the translation equivalent in L1 facilitates access to meaning for the new L2 words. However, the model assumes that the link between the L2 lexicon and the conceptual store might be strengthened with increased proficiency (Kroll & Stewart, 1994). That is, new second-language learners initially learn the lexical translations of foreign words, but a stronger direct access to semantics will eventually develop with sufficient L2 proficiency.

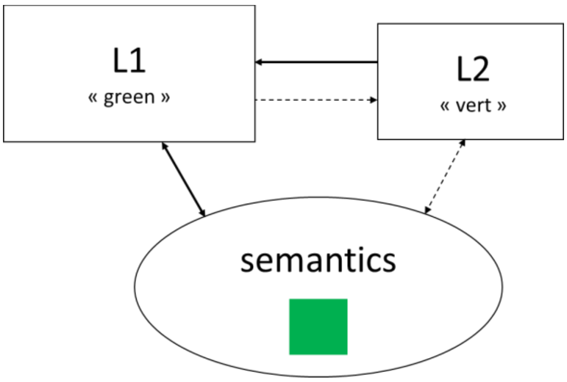


***Figure 4.*** Illustration of the Kroll and Stewart (1994) model

According to the Kroll and Stewart (1994) model, an English-French bilingual will, during early stages of L2 acquisition, associate the French word “vert” with its English translation equivalent, “green”. The English word “green” has privileged access to the meaning of the word, as a result of the strong word-to-concept connection for L1. In contrast, L2 words are not directly linked to the semantic store. That is, the semantic store is only accessed indirectly through lexical connections with L1. As L2 proficiency increases, the model hypothesizes that the connection between "vert" and the related concept will strengthen and the dependency on "green" as the L1 translation will diminish. This example demonstrates the developmental nature of the Kroll and Stewart (1994) model, by showing that the way in which lexical and conceptual information accessed changes as a function of the L2 proficiency level. Therefore, L2 processing requires the activation of the L1 equivalent until a certain level of L2 proficiency is reached.

The evidence that supports the Kroll and Stewart (1994) model originates mostly from translation experiments. According to the model, the forward translation (from L1 to L2) should be conceptually mediated, through the conceptual store since the lexical link in this direction is weak. On the other hand, translation in the backward direction (from L2 to L1) is lexically mediated, represented by the strong lexical link from L2 to L1. A group of fluent Dutch-French bilinguals performed the forward and backward translation task. Critically, two lists of stimuli were used: categorized lists that contained words arranged by semantic category, and randomized lists in which words were presented in a randomized order. As already mentioned, forward translation is assumed to be conceptually mediated and since the translation of the categorized list also involves conceptual processing, it should slow down the performance. This categorical interference effect results in slower translation of words presented in categorized lists relative to words presented in randomized lists. Backward translation represents the translation from L2 to L1, which was assumed to be unaffected by the semantics, and should, therefore facilitate performance in the randomized condition. Both hypotheses were supported: for the categorized list condition, forward translation took on average 1350 ms, as compared to 1150 ms for backward translation. Forward translation, hypothesized to be conceptually mediated, was slower for the categorized than for the randomized lists (i.e., impaired by the semantic manipulation). Slower performance was explained by activating the underlying concepts that slows down the translation process, as well as the less strong L1-L2 lexical link. For the randomized list condition, backward translation was significantly faster than the forward translation, reflecting the strong L2-L1 link and the absence of the concept activation evoked by the L2 words. Thus, the backward translation that is lexically mediated was unaffected by the semantic manipulation. The authors observed the translation asymmetry demonstrated by longer translation latencies in the L1 to L2 direction than vice versa. This asymmetry was observed in both proficient and less proficient bilinguals, although it was larger for the latter group of participants (Kroll & Stewart, 1994).

Another study which supports the Kroll and Stewart model (1994) investigated a transfer paradigm in relatively fluent English-Spanish bilinguals (Sholl et al., 1995). This paradigm was used to examine the relationship between picture naming and translation. Participants first named pictures in both L1 and L2, and then translated words from one language to the other in a subsequent task. Half of the words in the translation task corresponded to the concepts that had previously been named as pictures (this manipulation aimed to engage conceptual processing) and another half were new items. The Kroll and Stewart (1994) model predicts that the processing of the concepts in the picture-naming task would facilitate later translation of the same concepts, compared to the novel items. Picture naming produced reliable transfer to translation in the L1-L2 direction, but no transfer to translation in the L2-L1 direction. In other words, only the forward translation benefited from the prior conceptual processing. Backward translation seems to occur on the lexical level exclusively, without involving related concepts, which is in line with the assumptions of the model.

Asymmetries in translation process were further investigated in proficient bilinguals and in L2 learners. Relatively proficient and less proficient English-Spanish bilinguals performed a translation recognition task in which subjects had to respond whether the second word in the word pair is the correct translation of the first one. Of special interest were the words that were not correct translation equivalents, but related by lexical form (e.g. “man”-“hambre”; i.e., hunger) or by meaning (e.g. “man”-“mujer”; i.e., woman). They found greater interference of semantically related false translations relative to unrelated controls in a translation recognition task with highly proficient L2 subjects, whereas less proficient L2 subjects suffered more interference from form-related words. The results demonstrated the shift from reliance on word form in less proficient bilinguals to reliance on meaning (i.e., semantics) in high proficient bilinguals. This suggests that with the increased proficiency in L2, the ability to retrieve semantic information directly from the L2 word also increases. By relying on the direct L2 semantic representations, high proficient bilinguals may avoid lexical competitions between words that share lexical features across languages (Talamas et al., 1999).

A later study (Bowers & Kennison, 2011) replicated the Kroll and Stewart (1994) study, but also took into consideration the age at which words are acquired, normally referred to as *age-of-acquisition*. The authors hypothesized that the strength of the conceptual link between the L1 and the conceptual store is stronger for words learned early in life than for words learned later. The forward and backward translation paradigms were used. Half of the L1 words were learned early in childhood (early learned L1, e.g., sheep, mouth) and half were learned later in life (late learned L1, e.g., beaver, jaw). As in the original study, the L1 words were presented in both a semantically categorized condition (i.e., six semantic categories, such as animals, body parts, etc.) and a randomized condition. The study replicated the category interference effect observed in the Kroll and Stewart experiment for forward translation (L1-L2), but only in the case when the to-be-translated L1 words were learned early in life. When the L1 words learned later in life were translated into L2, the category interference effect had disappeared, as well as in the condition in which L2 words were translated into L1. Participants took longer to translate L1 words into L2 than they took to translate L2 words into L1 for late learned words and early learned words presented in the semantically categorized condition. However, participants translated as quickly from L1 to L2 as from L2 to L1 when early learned words were presented in random order.

***Critiques of the Kroll and Stewart (1994) model***

Although there are certain aspects of the Kroll and Stewart (1994) model that have been supported throughout the years, there are several shortcomings of this model. Some authors have criticized certain assumptions of the model (Duyck & Brysbaert, 2004; Pavlenko, 2009). The Kroll and Stewart (1994) model states that there are no strong conceptual links between L2 words and the conceptual system, at least at early stages of second language acquisition. As already discussed, the model assumes that the link between the L2 lexicon and the shared conceptual store may increase in strength as a function of L2 proficiency. However, the present model does not specify how these structures evolve over time and which mechanism leads to the strengthening of the lexical and conceptual links. In other words, the Kroll and Stewart (1994) model does not predict how lexical and conceptual systems develop as a result of learning (Heredia & Cieślicka, 2015).

Furthermore, Kroll and Stewart (1994) assumed that novel L2 words were acquired from lexical connections with their L1 counterpart. This model, therefore, does not take into consideration L1-L2 translation equivalents that do not share the same semantic representations across languages. Moreover, there are certain words that exist only in one language and, therefore, they cannot be linked to the lexical equivalent in another language (Pavlenko, 2009). Duyck and Brysbaert (2004) addressed this issue using a connectionist model. It assumes that connection weights between lexical and semantic levels may be influenced by two factors: semantic features of to-be-learned words and the number of connections the word has to other words. For instance, semantic weight should vary as a function of the amount of semantic overlap between two translation equivalents (Duyck & Brysbaert, 2004).

### *Semantic effects in L2 processing*

Another problem with the Kroll and Stewart (1994) model concerns semantic effects demonstrated in backward translations, suggesting that strong lexicosemantic links may exist for L2 words (Duyck & Brysbaert, 2004; La Heij et al., 1996). In a series of studies, La Heij and colleagues (1996) investigated whether backward translation (L2-L1) is semantically mediated. A group of unbalanced Dutch-English bilinguals performed reading and translating tasks on a set of congruent and incongruent Dutch (L1) and English (L2) colour words. As expected, reading was faster than translating. Interestingly, congruent colour words were translated faster than incongruent colour words, with the identical pattern in both translation directions. The authors assumed that incongruent colour words from both languages slowed down the translation process, suggesting that L2 words activate their meaning during translation. In another experiment, an English word (e.g., "turtle") was accompanied with a congruent picture (e.g., picture of a turtle) or completely unrelated picture (e.g., picture of a lighter). The semantic context effect (i.e., difference in translation latencies between unrelated and congruent condition) was larger in backward than in forward translation. This large effect observed in backward translation suggested that this translation direction is achieved though activation of target word's conceptual representations (La Heij et al., 1996).

The issue of semantic mediation in L2 processing was further examined in unbalanced bilinguals. Using the semantic Simon paradigm with a letter-case judgement task, Duyck and De Houwer (2008) investigated whether L2 words are potent enough to activate their meaning during low-level word processing. Dutch-English bilinguals had to classify the letter case of the target word using verbal labels. For instance, they had to say "animal" when the target word was written in capital letters or to say "occupation" when the target was written in lowercase letters. The letter-case judgement performance was measured on congruent trials when the verbal response matched the semantic category that the target word belonged to (e.g., saying “animal” to “lion”) and incongruent trials when the response mismatched the semantic category suggested by the target (e.g., saying “occupation” to “lion”). The results showed that both L1 and L2 target words yielded faster responses if the verbal target label (e.g., “animal”) matched the semantic category of the target (e.g., “lion”) than when it did not (e.g., lawyer). Although the meaning of the target was irrelevant for the task, a congruency effect was observed in both L1 and L2. The authors suggested that L2 words are potent enough to automatically activate semantic access during word processing through strong form-to-meaning mappings (Duyck & Brysbaert, 2004; Duyck & De Houwer, 2008).

Another evidence for conceptual mediation in backward translation comes from studies with numerical stimuli. The experiments were based on the *number magnitude effect*, an observation that the time needed to process numbers depends on the magnitude of the number. Duyck and Brysbaert (2004) examined whether this number magnitude effect occurs in bilinguals when they are asked to translate numbers in both forward (L1-L2) and backward (L2-L1) directions. They found that it takes longer to translate number-words that represent a larger quantity (e.g., “eight”) than number-words for a smaller quantity (e.g., “two”) in both L1-L2 and L2-L1 translation directions, regardless of word length or frequency. The number magnitude effect observed in bilinguals suggests that L2 words have strong semantic background. Importantly, these effects were observed with a set of novel number words that participants had acquired one hour before testing. This suggested that novel L2 words are not acquired through lexical connection with their L1 counterparts, but by early mapping with semantic representations (Duyck & Brysbaert, 2004).

### *Conclusions*

Numerous studies have called into question certain features of the Kroll and Stewart (1994) model of bilingual memory representation. The studies presented above used different paradigms to examine the presence of conceptual connections between L2 vocabulary and underlying semantic representations. Even though they mostly tested unbalanced bilinguals, it is plausible that their L2 proficiency was sufficiently high to produce semantic effects in foreign word translation. The line of research discussed below aimed to test whether conceptual connections occur for novel foreign words in recently trained bilinguals by using the Stroop paradigm. However, before turning to a chapter dedicated to foreign word learning and training, I discuss the role of the Stroop paradigm in investigating a bilingual interference effect and how the corresponding findings contribute to understanding bilingual memory organization.

## 3. Bilingual Stroop task

The Stroop task has been frequently used to study interference effects in bilinguals (Altarriba & Mathis, 1997; Dyer, 1971; Mägiste, 1984; Preston & Lambert, 1969; Schmidt et al., 2018; Tzelgov et al., 1990). It is known from previous studies that the congruency effect can be observed with both first language (L1) and second language (L2) distracter words. Interference in colour naming produced by distracting words increases as a function of its colour relatedness (Klein, 1964, see *Conflicts in the Stroop task* section). Consequently, incongruent colour words in a foreign language are assumed to cause much more interference to colour naming for persons with knowledge of that language relative to persons with no such knowledge (Dyer, 1971). For instance, an English-French bilingual will be impaired by both English and French incongruent colour words relative to an English monolingual who probably will not experience interference effect produced by French distracters. An interesting line of research investigated the magnitude of interference modulated in function of similarity between stimulus and response languages (Dyer, 1971; Preston & Lambert, 1969).

***Within (intralingual) and between (interlingual) language interference***

An important question in the bilingual Stroop literature is how Stroop interference is modulated by stimulus and response languages (i.e., the language of the stimulus words and the language in which participants need to respond, respectively). On the one hand, the language of the distracter word can match the language used for responding. For instance, colour naming of the distracter word “red” printed in blue produces within-language (intralingual) interference if the response language is English (i.e., the correct response is to say “blue”). On the other hand, the language of the distracter word can mismatch the response language. For example, English colour naming of the distracter word “rouge” (i.e., red in French) printed in blue results in between-language (interlingual) interference.

The interference effects described above are found for monolinguals (within-language) and bilinguals (between-language). For instance, one might wonder whether colour naming of distracter “red” printed in blue produces the same amount of interference as distracter “rouge” printed in blue for English-French bilinguals. Considering semantics, both “red” and “rouge” are assigned to the same concept that diverges from the one that represents the actual target (blue). For this reason, a similar amount of within-language and between-language interference might be expected. The magnitude of within- and between-language interference has been compared in numerous studies. In his review of Stroop literature, MacLeod (1991) reported that the between-language interference typically represents about 75% of within-language interference. For instance, if colour naming of the stimulus “red” printed in blue leads to 100 errors, then the presentation of the stimulus “rouge” printed in blue should result in 75 errors (MacLeod, 1991).

Several other studies supported this notion of larger within-language interference than between-language interference, conducted on Chinese-English and Japanese-English bilinguals (Fang et al., 1981), and in Turkish-English bilinguals (Kiyak, 1982). An analogous effect has been found for auditory stimuli in French-English bilinguals. Subjects were instructed to analyse the characteristics of a speaker’s voice while ignoring the meaning of the spoken words. The language of the stimulus words and the response language varied systematically. Subjects had to judge whether the pitch of the speaker’s voice was high or low while responding in the same or in the other language from the one in which stimulus word (i.e., “high”, “low”, “haute”, or “basse”) was presented. A larger interference effect was evidenced when stimulus and response languages matched relative to when they mismatched (Hamers & Lambert, 1972). Results supported the notion that bilinguals are unable to ignore the semantic characteristic of stimuli, manifested in increased latencies when words were presented instead of tones (i.e., a neutral condition).

***Cognate status***

Note however that the magnitude of between-language interference at least partially depends on the similarity between languages. *Cognates* are translation equivalents that share orthography and phonology across languages (e.g., “bleu” and “blue” in French and English, respectively). Apart for orthographic (i.e., spelling) similarities, there is also a compatibility in pronunciation between cognates. In contrast, non-cognates are dissimilar in terms of spelling and pronunciation (e.g., “vert” and “green” in French and English, respectively).

The effects of cognates have been investigated in translation studies. The asymmetry in translation latencies that should occur as a function of translation direction (forward vs. backward), as suggested by the Kroll and Stewart (1994) model, was not evidenced for cognates. When the word pairs are cognates (e.g., “blue”, “bleu”, and “blauw” in English, French, and Dutch, respectively), translation latencies are similar in both translation directions (de Groot et al., 1994). Forward translation from native Dutch words to their English (L2) equivalents was faster and more accurate for cognates than for non-cognates (de Groot, 1992).

Effects of cognates were also observed in foreign language acquisition. For instance, cognates are easier to learn relative to non-cognates regardless of other factors such as word frequency, congruency between learning and testing conditions, and learning method (Lotto & de Groot, 1998).

***Cognates and between-language interference***

Certain researchers measured within-language interference and compared it with the magnitude of between-language interference when cognates are used. For instance, in series of studies conducted by Preston and Lambert (1969), two languages spoken by bilinguals were used in all combinations of stimulus presentation and colour naming. For instance, a group of English-French and English-Hungarian balanced bilinguals performed a task in which they had to name the print colours of: a) English colour words in English, b) French or

Hungarian colour-words in French or Hungarian, respectively (within-language interference), c) French or Hungarian colour words in English, d) English colour-words in French or Hungarian (between-language interference), e) asterisks in English, and f) asterisks in French or Hungarian (control condition). Additionally, they compared colour naming performance when the colour words were similar in the two languages (i.e., cognates) and when they were dissimilar (i.e., non-cognates). All experimental combinations of stimulus and response languages resulted in large interference effects. However, the greatest interference was observed when the distracting and naming languages were the same (e.g., “blue” printed in red, named in English). Interference was almost as large when two languages were different, but the corresponding colour names were similar in the two languages (e.g., “bleu”, French for blue, printed in blue, named in English). Thus, they concluded that in certain cases the Stroop effect between languages might be as large as within language interference (i.e., up to 95% for English-French bilinguals). The amount of interference was much lower (i.e., 68% for English-Hungarian bilinguals) when colour word equivalents had different phonologic and orthographic features (i.e., non-cognates). To further elaborate findings on cognates and non-cognates, the following study (Experiment 2) tested German-English bilinguals in all four combinations for naming and interference. The between-language similarity was manipulated by using either a set of German-English cognates (i.e., “grün”, “rot”, “blau”, and “braun” as German colour words and “green”, “red”, “blue”, and “brown” as their English translations, respectively) or a set of non-cognates (i.e., “schwarz”, “gelb”, “rosa”, “lila”, translated as “black”, “yellow”, “pink”, and “purple” in English, respectively) across two languages. All conditions produced interference in colour identification, but interestingly, cognates produced an almost equivalent amount of within- and between-language interference. The amount of between-language interference was significantly reduced for dissimilar colour word translations. The authors concluded that the high interference that occurs when the to-be-named colour words are similar to their equivalents in the naming language results from a greater tendency to translate (Preston & Lambert, 1969).

An alternative explanation assumes that the similar sounding word automatically activates its counterpart in the naming language, without translation. Dyer (1971) found that interference between stimulus and response languages may be observed even in monolinguals (i.e., persons who should easily ignore the distracting stimuli written in a foreign language because they do not speak that language). A group of English monolinguals had to name the colour of colour word distracters written in English or in several other languages. As expected, colour naming was slowest for English distracters and speeded as foreign colour word distracters became less similar to their English counterparts. For instance, Spanish word "azul" is far removed from its English translation ("blue"), producing less interference than German ("blau") or French ("bleu") equivalents. The distracter words are claimed to activate the corresponding, similar sounding word in the naming language, concurrently ruling out the role of translation. Thus, even when the foreign words are unfamiliar, cognates produce interference due to their similarity with L1 colour words.

Although a between-language interference effect was observed in monolinguals, its magnitude is more pronounced for bilinguals. In monolinguals, the link between the foreign colour name and its incongruent colour name in the naming language is weak, thus they show little interference. In bilinguals, on the other hand, this associative link is strong, which produces much more interference. Apart from the observed interference effects for monolinguals, another interesting finding from Dyer (1971) confirmed that interference is reduced for bilinguals in a between-language condition than in the within-language condition (Dyer, 1971; Preston & Lambert, 1969).

### *Magnitude of interference in L1 and L2*

Apart from cognate status, the interference effect is generally larger for L1 words than that for L2 words. That is, independent of response language, L1 words produce larger overall congruency effects. One explanation argues that the size of the interference effect could be proportional to language proficiency. Mägiste (1985) conducted several experiments on Swedish monolinguals, German-Swedish bilinguals, and trilinguals with a range of different native languages. The interference effect was measured with a Stroop task. Greater within-language than between-language interference was observed in Swedish monolinguals and in participants who were born in Germany and kept a German language dominance in the initial period of their residence in Sweden (German-Swedish bilinguals). After years spent in Sweden, native German speakers mastered Swedish, which resulted in an equal magnitude of within- and between-language interference (Mägiste, 1984, 1985). It seems plausible that the magnitude of between-language interference might be influenced even by subtle differences in language dominance and language exposure. For instance, a monolingual daily environment combined with low second language skills may have enabled participants to inhibit their second language sufficiently enough to avoid producing interference with the first language (Marian & Spivey, 2003).

### *Conclusions*

Numerous studies have investigated interference effects produced by L1 and L2 words, by manipulating factors such as similarity between languages, cognate status, response modality, etc. However, as already mentioned, relatively little is known about the source of this interference in L2 with different proficiency levels. The main hypotheses of the present thesis concern the source of the interference effect in a recently trained L2. In the subsequent chapter, I discuss L2 word learning procedures and related relevant empirical findings.

## 4. Novel word learning and training

The learning of novel L2 words for concepts that are already known requires establishing new form-meaning connections (Barcroft, 2003; Potter et al., 1984). As previously discussed, Kroll and Stewart (1994) assumed that in the initial stages of language acquisition, foreign language learners might experience difficulties in matching a novel word's lexical form with its corresponding meaning (Kroll & Stewart, 1994). Even though this connection between an L2 word and concept seems to be weak in the initial stages of L2 acquisition (Talamas et al., 1999), it can strengthen as L2 proficiency develops. In experimental settings, researchers usually adopt different methods to establish and reinforce foreign language learning.

### *General approaches*

Foreign language learning studies usually employ different methods to associate novel word forms with corresponding meaning, followed by a testing phase in which the strength of the established associations is measured. In certain novel word learning studies, L2 words were associated with their native language counterparts. This learning method is known as the *word-association* paradigm (Altarriba & Basnight-Brown, 2012; Altarriba & Mathis, 1997; Hummel, 2010). In some other studies, novel words were presented together with corresponding definitions (Clay et al., 2007; Tamminen & Gaskell, 2013), or pictures representing the corresponding concept (Clay et al., 2007; Dobel et al., 2010; Lotto & de Groot, 1998; Webber, 1978; Yu & Smith, 2007). These learning methods aimed to connect words and related concepts in a rich context emphasizing the role of semantics.

A word learning phase is typically followed by different types of tasks that aim to reinforce and test novel word-to-concept connections. For instance, the tasks such as translation matching (Dobel et al., 2010), object naming (Lotto & de Groot, 1998), recognition (Altarriba & Basnight-Brown, 2012), semantic priming (Dobel et al., 2010; Tamminen & Gaskell, 2013), or a series of quizzes aimed to strengthen semantic connections between novel words and underlying representations (Altarriba & Mathis, 1997). Results revealed the novel word-concept links occur relatively quickly. However, priming tasks and other explicit, non-speeded tasks are susceptible to participants' strategic manipulations and could possibly mask the nature of these conceptual links. For instance, priming effects might reflect the associative relation between words (Shelton & Martin, 1992), rather than the semantic relation. Also, these measures do not clarify how automatic the novel word's meaning really is (Geukes et al., 2015), and it is not clear what amount of training is required for the development of automatic semantic processing.

***Stroop task and testing foreign word acquisition***

The Stroop task has been suggested as an efficient alternative that can rule out the abovementioned issues (MacLeod, 1991; Stroop, 1935). It tests certain components of automaticity in word meaning access (Moors & De Houwer, 2006). In particular, it allows one to examine whether novel word reading will automatically activate its meaning, as learned in learning/training phases. The Stroop task, which reveals the influence of automatic semantic activation, could serve as an excellent tool to examine whether, how fast, and how strongly novel words are mapped to their corresponding concepts. Surprisingly, not many studies administered novel word learning procedures and tested their efficiency in activating the semantic connection with the Stroop task.

Altarriba and Mathis (1997, Experiment 2) trained a group of English monolinguals with a set of Spanish colour words. In the first phase, monolinguals learned three sets of English-Spanish colour-word pairs that were presented both visually and auditorily. The L1-L2 connection was further strengthened through the series of quizzes. In one task, participants were asked to write the English (L1) word that corresponded to a given Spanish (L2) word.

Once done, the experimenter corrected any mistakes and allowed participants to look over the correct responses, therefore providing them feedback. In a following task, participants had to fill in one of a set of Spanish words that could fit into a simple close-ended English sentence. Subsequent testing required either a speeded recognition or a colour-naming response (i.e., Stroop task). In the speeded recognition task, participants had to determine as quickly and accurately as possible if the correct translation pairing or incorrect translation pairing was presented.

In the Stroop portion of the task, both an English monolingual group (that recently received training on Spanish words) and an English-Spanish bilingual group of participants (which did not need Spanish word training and had participated only in a Stroop portion of the experiment) took part by naming the colour of the word aloud in English. A within-language Stroop effect was indicated by slower responses in the English-incongruent than in the English-congruent condition, while the significant between-language Stroop effect was demonstrated by slower response times in the Spanish-incongruent than in the Spanish-congruent condition. In other words, it took longer to name colours in the incongruent than in the congruent condition when the response and the target were in different languages (i.e., Spanish target and English response). The results revealed that both groups produced significant within- and between-language Stroop interference, even though the effect was numerically larger for English (L1) words. However, these results suggest that even recently acquired incongruent distracters can slow down colour naming (Altarriba & Mathis, 1997).

Another study used a Stroop task to test the automaticity of recently formed word-concept connections. Twenty-five pronounceable nonwords served as novel words in the study. Ten of them had to be learned, and they were paired with German colour words via paired association, while the remaining 15 nonwords served as fillers. During the learning phase, each German word (e.g., “blau”, German for blue) was paired 24 times with one, to-be-associated pseudoword (e.g., “alep”, as a match trial) and once with each of remaining pseudowords (nonmatch trials). This so-called *statistical learning* procedure required from participants to indicate whether the two presented words belong together or not, with matching and mismatching word pairs presented equally often. Initially, participants were unable to be certain whether a word pair matched or not (i.e., their initial responses were based on guessing). However, during the task, the more frequent co-occurrence of some word pairs was supposed to help them to discriminate between matching and mismatching pairs. Novel word acquisition was tested in a manual version of the Stroop task either immediately after the learning phase or 24 hours later. The results revealed that when the novel words were intermixed with German colour words in the Stroop task, the Stroop effect was evidenced immediately after the learning phase. In contrast, if no German colour words were presented together with the recently learned pseudowords, the Stroop effect occurred only after 24 hours. The authors concluded that the automatic availability of the novel, to-be-learned words depends either on the supportive context provided by the familiar paired words or on sufficient time necessary for memory consolidation (Geukes et al., 2015).

Semantic learning of novel words was investigated in the picture-word interference (PWI) task. In this variant of the Stroop task, participants were slower and less accurate to name a target picture when it was presented with a semantically related written distracter, known as *picture-word interference (PWI) effect*. Even though the distracter word is supposed to be ignored, it automatically activates its meaning, and semantic competition slows down the target naming (see Risko et al., 2005). Picture naming is delayed by the simultaneous presentation of an unrelated distracter word compared with a nonword (i.e., a *general PWI*), as well as by a semantically related word relative to an unrelated distracter word (i.e., a *specific PWI* effect). Participants learned a set of 12 novel words (e.g., “kosla”) paired with an attributed description and picture (e.g., a bitter and spiky fruit). Descriptions and pictures were selected from one of three categories: fruits (e.g., grape, plum), vehicles (e.g., car, ship), and clothes (e.g., blouse, skirt). In the training phase, either a written description or a picture was presented on the screen with a novel word bellow it. Participants had to decide whether the novel word matched the description or picture. In the test phase, participants named aloud the line drawings of 12 familiar objects. They were instructed to name the drawing as quickly and as accurately as possible while ignoring the letter string distracter presented with it. The distracter was, depending on the condition, either a familiar word, a recently trained novel word, or an untrained novel word. The results for newly acquired distracters revealed that picture naming was delayed by semantically unrelated novel words relative to untrained novel words (i.e., general PWI effect). This result showed that the recently acquired semantic knowledge (i.e., information about the new words) can be processed automatically even after relatively little practice. Also, participants took longer to name pictures paired with semantically related novel words than to name those paired with unrelated novel words (i.e., specific PWI effect), but only a week after the training (Clay et al., 2007).

The studies presented above (Altarriba & Mathis, 1997; Clay et al., 2007; Geukes et al., 2015) investigated the presence of Stroop interference effects for recently learned words. Although interference was evidenced in the novel L2 words, the source of this effect was not clear. Another series of studies analogous to language learning investigated the source of this interference. Recently, Liefooghe and colleagues (2020) trained participants with nonwords via *conditional-discrimination training*. First, in a matching-to-sample task, participants were presented with a target (i.e., sample) stimulus (e.g., “plesk” or “klamf”) and two comparison stimuli (e.g., colour words, such as “red” and “green”) that they had to choose between. The relation between target and comparison stimuli was reinforced through feedback. For instance, when participants correctly match the comparison stimulus “red” with the sample stimulus “plesk”, and the comparison stimulus “green” with the sample stimulus “klamf”, the connection between two stimuli has been reinforced. A following matching-to-sample phase had no feedback and it required matching not only previously-reinforced word pairs (e.g., "red"-"plesk"), but also their reversed connections (i.e., when the sample stimulus is "red", and a comparison stimulus is "plesk"). This training was further extended to new contingencies. Novel nonwords "smelk" and "gilpt" have been reinforced after correct matching with "plesk" and "klamf", respectively. Therefore, two equivalence classes (i.e., the first one containing "red", "plesk", and "smelk" and the second one containing "green", "klamf", and "gilpt") were formed. In a subsequent test phase, apart from previously reinforced connections (i.e., "red"-"plesk", "plesk"-"red", "green"-"klamf", "klamf"-"green"), participants were asked to match the latest introduced nonwords (i.e., "smelk" and "gilpt") with colour words. For instance, the comparison stimuli "smelk" was more likely to be selected when the sample stimulus was "red" although they had never been presented together. Similarly, the comparison stimuli "gilpt" was more likely to be selected when the sample stimulus was "green". This procedure allowed distinguishing three types of stimuli: colour words (i.e., "red" and "green"), reinforced associates (i.e., directly linked to colour words; "plesk" and "klamf") and derived associates (i.e., linked to reinforced associates but never directly linked to colour words; "smelk" and "gilpt"). After training, participants completed a 2-to-1 Stroop task with the colour words and their "associates" as distracters. Of particular interest were the effects of reinforced and derived associates on colour identification. That is, whether "smelk" as derived associate for "red" could speed colour naming when presented in the congruent (i.e., "smelk" printed in red) relative to incongruent (i.e., "smelk" printed in green) condition. Results showed that there was a response conflict effect observed for directly reinforced and derived associates. The interpretation of the results is based on the lexical processing demonstrated though translation of associates into their corresponding colour word. In other words, directly reinforced associates (e.g., "plesk") and derived associates (e.g., “smelk”) were automatically translated to their colour equivalent (e.g., “red”), activating the response linked to this colour. The results are not surprising since the connections were formed only between lexical representations of colour words and nonwords, with no semantic component. These associative-learning studies are, of course, only analogous to language learning, but suggest potential influences of stimulus and response conflict for newly acquired “words”.

Besides the presented learning studies, the present thesis has been directly inspired by one study on interference effects in bilinguals that more directly aimed to explore the source of interference for L2 words. Schmidt and colleagues (2018) administered a standard colour-identification Stroop task in Dutch-French bilinguals with Dutch (L1) and French (L2) distracting colour words. As an important note concerning the overall French (L2) language skills in their sample: participants were generally familiar with French but had only weak French skills. They used the 2-to-1 keypress mapping procedure to separate stimulus conflict and response conflict effects in both L1 and L2. As expected, first language words produced both stimulus and response conflict. This is not surprising if we consider the strong lexical and conceptual connections established for well-known L1 words. However, a particular question of interest of Schmidt and colleagues (2018) was whether the same pattern of results applies for the second language words. The results revealed both stimulus and response conflict for a second language. The stimulus conflict effect (i.e., difference in performance between identity and same response trials) is consistent with notion that L2 distracters are potent enough to activate the underlying semantic representation and interfere with the semantic identification of the relevant ink colour (Altarriba & Mathis, 1997). Furthermore, the response conflict effect (i.e., difference in performance between same and different response trials) suggests that the L2 words retrieve the response associated with its L1 translation, thereby producing the response conflict effect. Results suggested that there is a certain degree of similarity in semantic and response processing between L1 and L2 words (Schmidt et al., 2018). These findings contradict the assumptions of Kroll and Stewart (1994) according to which only response conflict effect should occur for L2 words. However, it is possible that the participants were too fluent in their L2 (French), producing a pattern that deviates from the predictions of the Kroll and Stewart model (1994). To further investigate this line of research, I opted to use less proficient and recently trained L2 in the subsequent experiments.

***The present work***

The crucial question of the present thesis is where conflict occurs within the cognitive system for weakly-proficient second language speakers and for recently trained L2 words. Therefore, I aimed to investigate whether L2 distracters evoke both stimulus and response conflict, or only one of the two. This could have important implications for models of bilingual language cognition. Since much debate has centred on the connection between bilingual lexicons and semantics, as in the Kroll and Stewart (1994) model (for more details, see Chapter 2), I believe that the present work will provide new insight to this question. I start this reasoning from the perspective of the Kroll and Stewart (1994) account. As already discussed, according to this account, L2 words are strongly connected to their L1 lexical representations (i.e., they are learned as translations of their corresponding, well-known L1 translations), but have relatively weak connection with their semantic concepts. For instance, a native English speaker that studies French will associate the novel L2 word “vert” with its L1 equivalent (i.e., “green”) at the lexical level, but the access of the novel word to semantic representation is assumed to be limited at early stages of L2 acquisition. This can serve as a starting point for discussing the source of conflict within the lexico-semantic system: at the level of semantics, at the level of responses, or both.

Three possible patterns of results might occur. First, both stimulus and response conflict effect will emerge for L2 words, similarly as for L1 words. In this case, same response trials should be slower than identity trials (i.e., stimulus conflict) but faster than different response trials (i.e., response conflict). This pattern would imply sufficiently strong lexical and conceptual connections for L2 words, similar as the ones of L1 words (Schmidt et al., 2018). On the other hand, the magnitude of conflict for L2 is smaller than the one for L1 (Mägiste, 1985), which suggests that one (or both) components are decreased for L2 words, but it could be that L2 words will still influence both semantic and response processing.

The second possible result is that the L2 words will produce stimulus conflict exclusively. This assumes that L2 words are directly linked to semantics, even when the overall L2 proficiency is not high. The literature on foreign language learning provides empirical evidence (Altarriba & Basnight-Brown, 2012; Altarriba & Mathis, 1997) for direct links between L2 words and semantics, even at early stages of L2 learning. Following the given example, the L2 word "vert" printed in red might interfere with colour identification (i.e., slower responding to name "red"). The presentation of foreign word "vert" for a native English speaker will activate the semantic representation for green, while the relevant stimulus dimension (i.e., ink colour) will activate corresponding semantic representation for red, producing conflict in semantics (Schmidt et al., 2013). However, according to this account, the L2 word will not interfere with response selection, since "vert" is unable to retrieve the response linked to green. In this case, L2 words will produce slower colour naming on same response trials than identity trials (i.e., stimulus conflict), with no difference in response times between same response and different response trials (i.e., no response conflict).

The third possibility is that L2 words will produce exclusively a response conflict effect. That is, an incongruent L2 stimulus (e.g., distracter "vert" printed in red) will interfere with response selection if it is potent enough to retrieve a corresponding response (i.e., press the key that corresponds to "green" or say "green" for manual and vocal variant of the task, respectively). For this to occur, “vert” has to be directly translated to its L1 translation (i.e., “green”), that will activate the corresponding response alternative, but without interfering with identification of the stimulus colour (i.e., “red”) itself (Kroll & Stewart, 1994; Liefooghe et al., 2020). If this assumption is correct, L2 distracters should slow down response latencies on different response trials relative to same response trials (i.e., response conflict), but no difference between identity and same response trials is expected (i.e., no stimulus conflict).

This research question was investigated in the present series of studies on weakly spoken (Experiment 1) and recently trained (Experiment 2-6) second languages. The presence of either or both stimulus and response conflict effects for L2 colour words might depend on the word type. As already discussed, cognates are translation equivalents that share orthographic and phonological features in both languages (e.g., “bleu” in French and “blue” in English). In contrast, non-cognates are dissimilar across languages with little or no overlap in pronunciation and spelling (e.g., “vert” in French and “green” in English). To avoid similar influences of cognates on foreign language acquisition that are less interesting for the purpose of the present thesis, French-English (Experiment 1), French-Croatian (Experiments 2, 3, 5, and 6), and English-Croatian (Experiment 4) non-cognates were used.

## Experiment 1

Experiment 1 replicates a study conducted by Schmidt and colleagues (2018) on unbalanced Dutch-French bilinguals (see Introduction). This study revealed both stimulus and response conflict effects for a second language (French), as for a first language (Dutch). Interestingly, no significant difference in magnitude of these effects was observed across languages in response times, although the effects were numerically smaller for L2 words. This contradicted the hypothesis that L2 colour words produce stimulus conflict effect only, just as L1 colour associates do (Schmidt & Cheesman, 2005). The results are also inconsistent with the assumption that L2 words are not potent enough to bias a potential response. However, the assumption that L2 colour words only influence response selection was also not confirmed. Therefore, the results of Schmidt and colleagues (2018) do not support the notion that L2 words can only activate response associated with their L1 translation through lexical link (that should result in response conflict), but not activate semantic representations (to produce stimulus conflict), as assumed by Kroll and Stewart (1994). To sum up, the study on unbalanced Dutch-French bilinguals revealed that L2 words can both: 1) activate their semantic representations, and 2) bias a potential response, similarly as L1 words (Schmidt et al., 2018).

Concerning the results reported by Schmidt and colleagues (2018), some limitations of this study should be pointed out. For instance, the presence of both stimulus and response conflict effect for L2 words is not in line with some models of bilingual memory organization. As already mentioned, this contradicts the Kroll and Stewart (1994) model which assumes L2 response conflict exclusively. However, it seems plausible that the participants were too fluent in their L2 (Schmidt et al., 2018), thus producing a pattern of results that deviate from the assumptions of the model. To further explore this idea, we opted to use a less proficient L2 in Experiment 1. Thus, the crucial difference in the present experiment relative to original one conducted by Schmidt and colleagues (2018) is lower L2 proficiency in the sample.

## Method

### *Participants*

A total of 85 University of Burgundy undergraduates (72 women, 13 men) participated in the study. Participants were recruited by signing up on a sheet posted on the psychology department board. All participants had normal or corrected-to-normal colour vision and spoke French as a first language. They received course credit in exchange for participation.

### *Apparatus and materials*

Stimuli were presented on a standard 15'' PC laptop. Stimulus presentation and response collection were controlled by E-prime 2.0 software. Responses were made on a standard AZERTY keyboard, with the "F" (left) and "J" (right) keys. Prior to the computer portion of the experiment, participants were given a pen-and-paper survey to fill out. The first part of the survey was the LexTALE (Lemhöfer & Broersma, 2012), with instructions translated into French. Within this test, 60 English-looking words are presented. About 2/3 of the presented words are actual English words (e.g., "moonlit"), whereas the remaining 1/3 are not (e.g., "plaudate"). Participants were informed to select the words that they are certain are actual English words. Correct answers were rewarded with one point, and incorrect trials were penalized by two points. The questionnaire also asked for gender, native language, years of English study in school, and a self-rating of English knowledge on a scale from 0 ("almost none") to 10 ("perfect"). After this, a subset of questions from the French version of the Language Experience and Proficiency Questionnaire – LEAP-Q (Marian et al., 2007) were appended. The first three questions from the Questionnaire were retained, which asked, respectively, for a list of languages in order of dominance, a list of languages in order of acquisition, and the percentage with which the participant used each of their spoken languages in the recent period. Also retained from the LEAP-Q were two boxes, one for French and another for English, asking for the age that the participant began acquiring the language, became fluent in the language, began learning to read in the language, and became fluent in reading the language. The purpose of these questions was to assure participants had correct language dominance. These metrics were also correlated with the observed congruency effects. Finally, participants were asked to give the English translations of the four French colour words used in the experiment. The purpose of this question was to see how familiar the stimuli were to participants and to assure they knew the correct translation of each of the colour words.

***Design***

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

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

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

### *Procedure*

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

## Results

### *Language demographic*

For almost all participants, French was the first language in order of dominance and order of acquisition (98.82%). Participants mostly indicated English (68.24%), Spanish (11.76%), Turkish (3.53%), Portuguese (3.53%), and Arabic (3.53%) as second languages in order of dominance. Similarly, English (69.41%), Spanish (7.06%), German (4.71%), Turkish (3.53%), Portuguese (3.53%), Italian (3.53%), and Arabic (3.53%) were rated as the most frequent languages in order of acquisition. Participants mainly use French in their everyday life (*MEAN* = 78.26%, *SE* = 1.722), and they have relatively little exposure to English (*MEAN* = 13.69%, *SE* = 1.139). Average French and English language metric scores are presented in Table 1. Brief inspection of the age of gaining English language skills (speaking, reading) suggest they started quite late (9-15.5 years). Despite learning English for more than 9 years on average, participants self-rated their English proficiency relatively moderately (5.50 on 1-10 scale), and the objective English vocabulary knowledge scores (*MEAN* = 67.01, *SE* = .948) were quite low. Participants were mostly familiar with English colour words used in the

Stroop task. The words “yellow” and “green” were correctly translated by 96.5% participants and word “pink” by 92%. Only half of participants in the sample correctly translated word “silver”.

***Table 1.*** Experiment 1 - Mean language scores with standard errors

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

***Stroop task***

Mean correct response time and percentage error data of the Stroop task were analysed. For error percentages, any responses above 2000 ms were considered spoiled trials and were excluded from analysis. As a supplementary analysis, in this and all subsequent experiments we included Bayesian statistics using the standard noninformative Cauchy prior in JASP (Marsman & Wagenmakers, 2017) with a default width of 0.707. Bayes factors with values higher than 100 (i.e., $BF_{10} > 100$) represent extreme evidence for alternative hypothesis.

**Response time.** To analyse response times, we conducted a congruency (identity vs. same response vs. different response) by language (French vs. English) within-subjects repeated measures ANOVA. The correct response time data are shown in Figure 5. There was a main effect of congruency, $F(2,168) = 36.48$, $p < .001$, $MSE = 1581.64$, $\eta_p^2 = .30$, $BF_{10} > 100$. However, the main effect of language was not significant, $F(1,84) = .81$, $p > .05$, $MSE = 1657.54$, $\eta_p^2 = .01$, $BF_{10} = .139$. The interaction between congruency and language was only marginally significant, $F(2,168) = 2.67$, $p = .07$, $MSE = 1637.75$, $\eta_p^2 = .03$, $BF_{10} = 46.25$.

Even though the interaction between congruency and language was only marginal, we are inherently interested in knowing the separate results for each language individually (in this and all subsequent experiments). Thus, the comparison of response times between different types of trials was conducted separately for French and English words. For French colour words, we found significant stimulus conflict (same response – identity), $t(84) = 4.77$, $p < .001$, $MEAN_{diff} = -34.06$, $SE_{diff} = 7.14$, *Cohen's d* = -.52, $BF_{10} > 100$, and significant overall Stroop interference (different response – identity), t(84) = 7.19, p < .001, $MEAN_{diff} = -42.23$, $SE_{diff} = 5.87$, *Cohen's d* = -.78, $BF_{10} > 100$. Surprisingly, the response conflict effect (different response – same response) failed to reach significance, $t(84) = 1.25$, $p = .21$, $MEAN_{diff} = -8.17$, $SE_{diff} = 6.52$, *Cohen's d* = -.14, $BF_{10} = .254$. For English colour words, both stimulus conflict, $t(84) = 2.28$, $p < .05$, $MEAN_{diff} = -13.89$, $SE_{diff} = 6.10$, *Cohen's d* = -.25, $BF_{10} = 1.375$, and response conflict, $t(84) = 3.01$, $p < .01$, $MEAN_{diff} = -16.31$, $SE_{diff} = 5.41$, *Cohen's d* = -.33, $BF_{10} = 7.853$, were significant. The overall Stroop effect was significant; t(84) = 5.28, p < .001, $MEAN_{diff} = -30.21$, $SE_{diff} = 5.72$, *Cohen's d* = -.57, $BF_{10} > 100$.

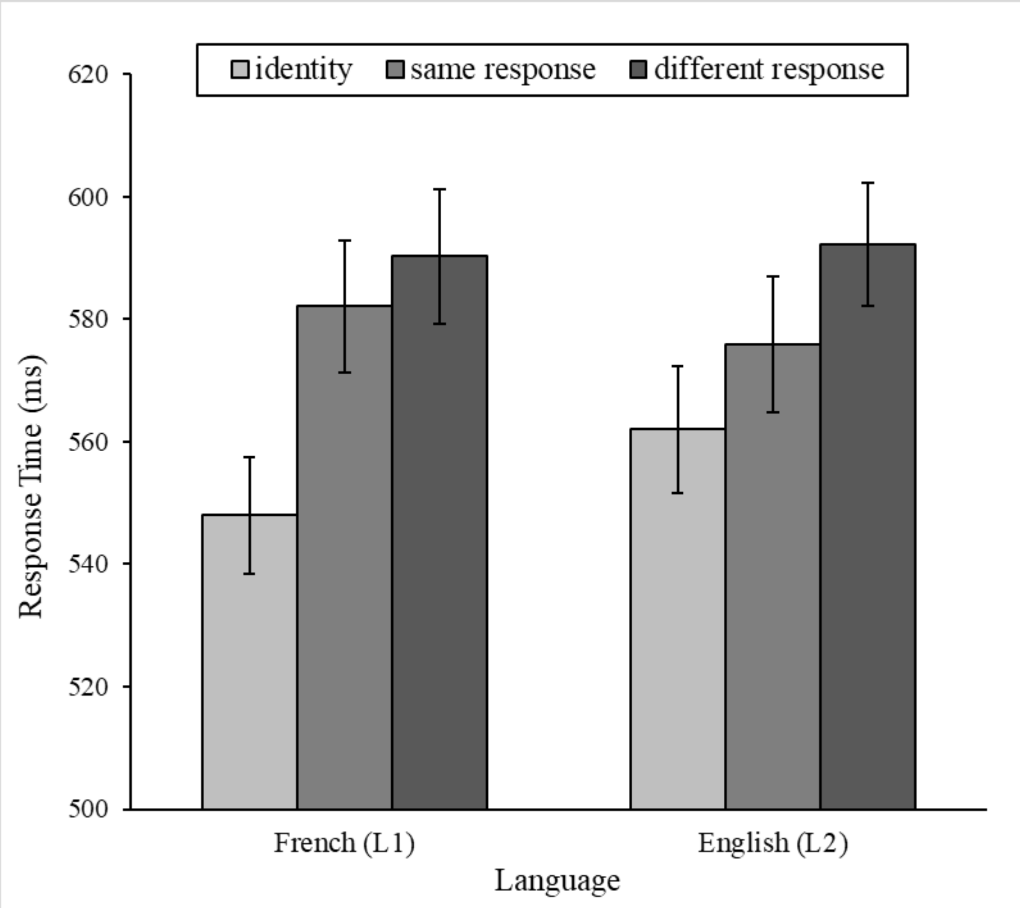


***Figure 5.*** Experiment 1 - Response times with standard errors for French and English colour words in the Stroop task

The magnitude of the stimulus conflict effect was larger for French than for English words, $t(84) = 2.20$, $p < .05$, $MEAN_{diff} = 20.17$, $SE_{diff} = 9.17$, *Cohen's d* = .24, $BF_{10} = 1.173$, but there was no difference in the magnitude of the response conflict effect across languages, $t(84) = .88$, $p > .05$, $MEAN_{diff} = -8.14$, $SE_{diff} = 9.27$, *Cohen's d* = -.09, $BF_{10} = .174$.

**Percentage error.** As for the response time data, we again conducted a congruency (identity vs. same response vs. different response) by language (French vs. English) within-subjects repeated measures ANOVA. The mean percentage error data are presented in Figure 6. The main effect of congruency was significant, $F(2,168) = 12.45$, $p < .001$, $MSE = 19.70$, $\eta_p^2 = .13$, $BF_{10} > 100$, but there was no main effect of language, $F(1,84) = .14$, $p > .05$, $MSE = 16.57$, $\eta_p^2 = .002$, $BF_{10} = .104$. The interaction between congruency and language was not significant, $F(2,168) = 1.71$, $p > .05$, $MSE = 15.08$, $\eta_p^2 = .02$, $BF_{10} = .18$.

Further comparisons between different types of trial were conducted on each language separately. For French colour words, there was no stimulus conflict effect, $t(84) = .32$, $p > .05$, $MEAN_{diff} = .19$, $SE_{diff} = .58$, *Cohen's d* = .03, $BF_{10} = .123$, but there was a significant response conflict effect, $t(84) = 4.51$, $p < .001$, $MEAN_{diff} = -2.74$, $SE_{diff} = .61$, *Cohen's d* = -.49, $BF_{10} > 100$. Similarly for English colour words, the stimulus conflict effect was not significant, $t(84) = 1.17$, $p > .05$, $MEAN_{diff} = .83$, $SEdiff = .71$, *Cohen's d* = .13, $BF_{10} = .276$, but the response conflict effect was significant, $t(84) = 2.77$, $p = .01$, $MEAN_{diff} = -1.83$, $SE_{diff} = .66$, *Cohen's d* = -.30, $BF_{10} > 100$.

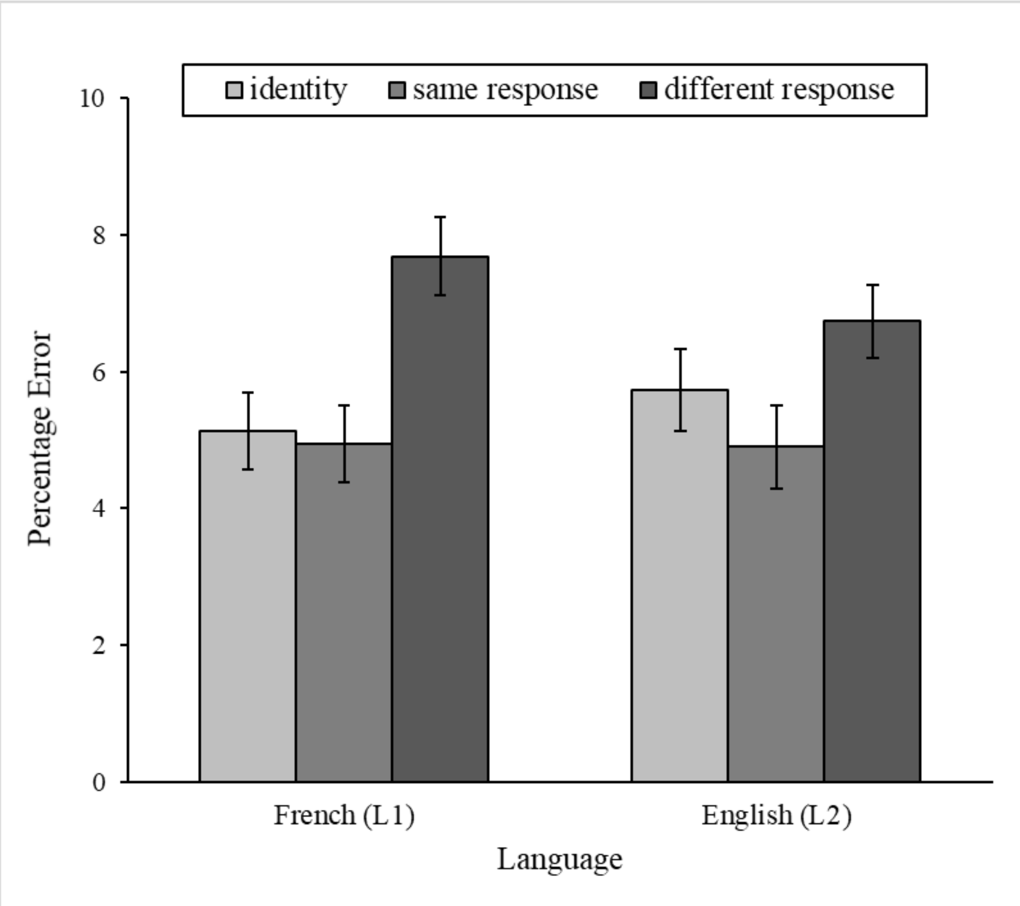


***Figure 6.*** Experiment 1 - Percentage errors with standard errors for French and English colour words in the Stroop task

There was no evidence for any differences in the magnitude of the stimulus conflict effect, $t(84) = .71$, $p > .05$, $MEAN_{diff} = .65$, $SE_{diff} = .90$, *Cohen's d* = .08, $BF_{10} = .153$, or the response conflict effect $t(84) = 1.16$, $p > .05$, $MEAN_{diff} = .90$, $SE_{diff} = .78$, *Cohen's d* = .13, $BF_{10} = .229$ across languages.

### *Correlations*

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

***Table 2.*** Experiment 1 - Correlations of language demographic variables with conflict effects

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

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

**Discussion**

Experiment 1 investigated the source of congruency effects in a weakly spoken second language. Conceptually, this experiment was a replication of the study that uses a more fluent second language (i.e., French) in Dutch-French bilinguals (Schmidt et al., 2018). Therefore, Experiment 1 adopted the same experimental design but, importantly, tested the extent to which the congruency effects occur with lower L2 proficiency.

Similarly as in the original study (Schmidt et al., 2018), Experiment 1 evidenced both stimulus and response conflict effects for a second language (i.e., English). Results for L2 words are in contrast with the hypothesis that second language words produce exclusively a stimulus conflict effect (Glaser & Glaser, 1989; MacKinnon et al., 1985) or exclusively a response conflict effect (Klein, 1964; Sharma & McKenna, 1998). These results challenge the notion that foreign language colour words do not influence response processing. A significant difference between same response and different response English colour word trials indicates that incongruent L2 words do bias a potential response. The L2 words retrieve the response associated with its L1 equivalent (Kroll & Stewart, 1994). Additionally, the results support the notion that L2 words are potent enough to activate the corresponding semantic concept. A significant stimulus conflict effect observed for L2 colour words implies that foreign language words can tap into semantics without the lexical link with L1 (Duyck & Brysbaert, 2004).

Unexpectedly, only the stimulus conflict effect, but not the response conflict effect was significant in response latencies for the first language (i.e., French). This is almost certainly a Type 2 error given that response conflict has been observed repeatedly in L1 (Augustinova et al., 2015; De Houwer, 2003; Schmidt et al., 2018; Zhao et al., 2015). Although statistically non-significant, the L1 response conflict effect was still trending in the correct direction with slower response latencies for different response trials relative to same response trials. Moreover, this response conflict effect was large and robust in the errors. This might suggest a speed-accuracy trade off (or simply a Type 2 error in response time, as previously noted). Regardless of that, our main hypotheses were about L2 since L1 has already been studied repeatedly.

There was a sizeable difference in the stimulus conflict effect for response latencies observed across languages, with a larger effect for L1 words. This is in line with previous studies that evidenced larger interference produced by L2 relative to L1. However, Mägiste (1984) claimed that the amount of conflict is influenced by level of proficiency. Smaller effects observed for L2 words could be due to relatively low (both subjective and objective) English proficiency in our sample. No difference was observed in the magnitude of response conflict effect across languages.

To make sure participants had the desired linguistic background (i.e., weak L2 proficiency), we administered the LexTALE vocabulary test, self-report measures, and a LEAP-Q set of language demographic questions. The aim of these measures was to confirm weak English language skills within the sample. All the applied measures revealed a relatively poor L2 proficiency in the sample, weaker than the L2 proficiency in the original study (Schmidt et al., 2018). Therefore, we can confirm that the English proficiency in our French sample was lower that the French proficiency in the Dutch sample observed by Schmidt and colleagues (2018). Despite that, the observed results are quite similar to the results of Dutch-French sample. In both studies, L2 colour words produced the same pattern of results as L1 colour words. These results suggest that there is a certain similarity in semantic and response processing between L1 and L2. That is, English colour word (e.g., "green") that are orthographically and phonologically different from their native French equivalents (e.g., "vert") automatically interfered with stimulus and response processing.

One limitation of Experiment 1 is that French participants were certainly familiar with the English colour words used in this study. This is supported by relatively high accuracy of correct L2 colour word translations observed in the questionnaire portion of experiment. Colour words are usually studied in early phases of foreign language acquisition (Nikolov & Mihaljević Djigunović, 2011). Consequentially, those words might be well integrated into L2 vocabulary. It could be that during very early language acquisition only stimulus or only response conflict are present, but that our participants were sufficiently familiar with the colour words used in our study to produce both. This potential caveat was addressed by training participants with completely unfamiliar colour words in the following studies.

## Experiment 2

Experiment 2 had the same objective of investigating the source of the interference effect in L2, similar to Experiment 1. As already discussed, it is possible that the participants in Experiment 1 were more fluent than the objective and subjective L2 proficiency measures suggested, producing therefore contrasting results to Kroll and Stewart (1994). Thus, the next step was to exclude the possibility that participants are too fluent in L2 and to further test the assumptions of Kroll and Stewart (1994) with people who have just learned a set of words in a second language. According to their model, second language words could be only weakly connected to semantics. The link between novel L2 words and their semantic representations should develop as a function of L2 proficiency. This model assumes that at early stages of L2 acquisition, only a lexical link between a novel word and its first language equivalent is formed. In other words, a second language word is memorized as a direct translation of its first language counterpart. To test these hypotheses of Kroll and Stewart (1994), we introduced words from a novel obscure language (i.e., Croatian) and assured that our participants have no previous knowledge of it. According to Kroll and Stewart (1994), novice L2 learners without previous L2 knowledge will not be affected by L2 conceptual interference.

By integrating a learning portion in the present experiment, we could observe the source of interference that might occur for novel language words, as well as control the amount of training needed for this interference to occur. In the initial phase of Experiment 2, native French speakers learned a novel Croatian word (e.g., “zelena”) as a translation of each corresponding French word (e.g., “vert”; i.e., French for “green”). A novel word was also associated with the perceptual representation of the corresponding to-be-learned colour concept (i.e., “zelena” printed in green). The Kroll and Stewart model (1994) therefore assumes that a lexical link is established between “zelena” and “vert”, without any deeper semantic processing. However, connections between L2 words and underlying semantic concepts are found to emerge eventually as L2 proficiency increases (Sholl et al., 1995; Talamas et al., 1999). This implies building a direct conceptual link between, for example, “zelena” and the semantic concept of the colour green when sufficient L2 proficiency has been reached. This conceptual link is considered as independent of the lexical link. On the other hand, the model neither defines the features of the mechanism that strengthen lexical and conceptual links nor does it specify how these links evolve over time (Heredia & Cieślicka, 2015). To train our participants with novel L2 words, we adopted a short and simple training procedure (up to 10 minutes). Participants performed on a set of training trials in which they matched either French colour words with their Croatian counterparts or Croatian colour words with their French translations.

## Method

### *Participants*

A total of 116 University of Burgundy undergraduates (100 women, 16 men) participated in the experiment. The selection was based on the following criteria: participants needed to speak French as a native language, have normal or corrected-to-normal colour vision, and have no prior knowledge of the Croatian (or a similar) language. As required, none of participants spoke Croatian or a similar language, and they had not studied Croatian in school. More demographic information about the sample can be found in the Results section. They received course credit for their participation in the experiment.

### *Apparatus and materials*

As in Experiment 1, stimuli were presented on a standard 15’’ PC laptop. Stimulus presentation and response collection were controlled by E-prime 2.0 software. Prior to the experimental portion of the study, participants were given a pen-and-paper survey to fill out. Several questions were retained from French version of the Language Experience and

Proficiency Questionnaire – LEAP-Q (Marian et al., 2007), such as the ones concerning a list of languages in order of dominance, a list of languages in order of acquisition, and the percentage with which the participant used each of their spoken languages in the recent period. Also retained from the LEAP-Q was a box for French, asking for the age that the participant began acquiring the language, became fluent in the language, began learning to read in the language, and became fluent in reading the language. Additionally, the questionnaire asked for gender and native language. To ensure participants had a target language background that excludes the Croatian language, we asked for years of Croatian study in school and a self-rating of Croatian knowledge on a scale from 0 ("almost none") to 10 ("perfect"). As in Experiment 1, participants were asked to give the French translations of the four Croatian colour words (i.e., "crvena", "plava", "zelena", "siva") used in the experiment. We expected no participants would be able to translate these words correctly.

***Design***

After the survey, participants proceeded to the computer portion of the experiment. It contained three phases: learning phase, training phase, and Stroop task phase.

**Learning phase.** In the learning phase, participants were presented with the Croatian and French colour words for "red", "blue", "green", and "gray" (Croatian/French: "crvena/rouge", "plava/bleu", "zelena/vert", and "siva/gris", respectively). These words were non-cognates with a mean word length of 4.75. The colour words from both languages were chosen based on the similarity in word length and absence of special characters common for Croatian language (e.g., "ž", "đ", "č") which appear in some other Croatian colour words (e.g., "žuta/yellow", "smeđa/brown", etc.). The corresponding print colours were red (255,0,0), blue (0,0,255), green (0,128,0), and grey (128,128,128), corresponding to "red", "blue", "green", and "grey" in the standard E-prime/HTML colour palette. The learning phase consisted of one block of 4 word pairs presented 4 times each, in randomized order.

Participants were asked to pay attention to word pairs and try to memorize them. A schema of a learning trial is presented in Figure 7.

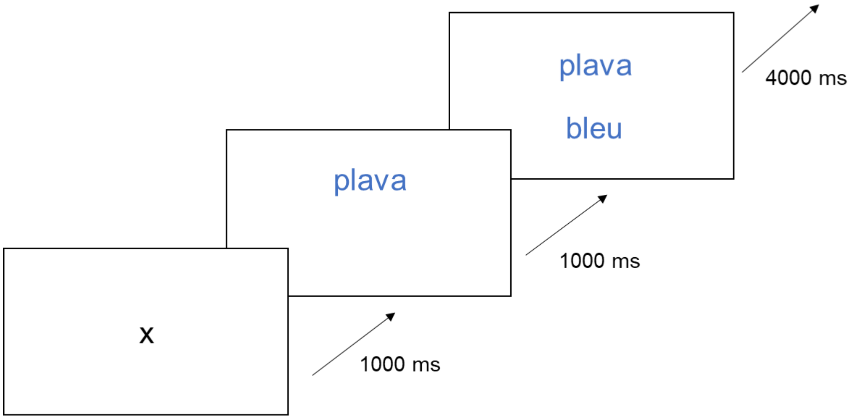


***Figure 7.*** Example of a trial in the learning phase

**Training phase.** In the following training phase, participants were presented with either a Croatian or French colour word printed in black. Below the target word, there were four labels printed in their corresponding colour (e.g., “zelena” or “vert” as potential responses were printed in green) with possible answers in the other language. The task was to choose the label with the accurate translation by pressing a proper key. Each of the eight colour words was presented 4 times as a target. A schema of the two types of training trials (French target with Croatian labels and Croatian target with French labels, respectively) is presented in Figure 8.

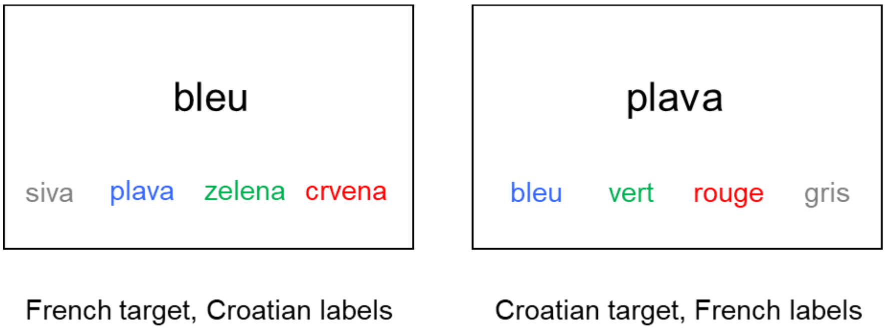


***Figure 8.*** Experiment 2 - Types of trials in the training phase

**Stroop task.** In the third, Stroop task phase of the experiment, participants completed the Stroop task. As in Experiment 1, we again utilized the 2-to-1 mapping procedure. The key mapping manipulation allow for two within factors; distracter language (Croatian vs. French) and congruency (identity/same response/different response). The structure of Stroop portion of the experiment was identical to that in Experiment 1.

### *Procedure*

Each trial in the *learning phase* started with a fixation "+" for 1000 ms. Next, a Croatian word appeared in the centre of the screen for 1000 ms. While the Croatian word remained on the screen, its French translation appeared below it. Both words remained on the screen for 4000 ms.

In each trial in the *training phase*, either a Croatian or French colour word printed in black appeared in the centre of the screen. Below it, 4 labels with possible answers printed in the corresponding colour were given together with the corresponding key ("d", "f", "k", and "l"). For instance, Croatian word "plava" was followed by 4 possible answers in French ("rouge", "bleu", "vert", or "gris"). The order of the labels was completely randomized on a trial-by-trial basis. The word was presented until a response was registered or 3000 ms elapsed. The next trial began immediately following a correct response. If the participant made an error or failed to respond in 3000 ms, the message "Erreur" ("Incorrect/Error") or "Trop lent" ("Too slow"), respectively, appeared in black for 1000 ms before the next trial. The time frames in the *Stroop task* phase remained unchanged relative to Experiment 1.

## Results

### *Language demographic*

For the majority of participants, French was the first language in order of dominance (99.1%) and acquisition (98.3%). As a second language in order of dominance, participants had mostly indicated English (74.14%), Spanish (20.69%), and German (1.72%). Similarly, most of them had acquired English (80.17%), Spanish (7.76%), German (4.31%), and Italian (2.59%) as a second language. Their estimated daily exposure rate to French language was

80.96% (SE = 1.406). Mean French language scores are presented in Table 3. None of the participants studied Croatian in school or spoke a similar language. All of them self-rated their Croatian knowledge as 1 (“not at all”) and were not able to translate the given Croatian words.

***Table 3.*** Experiment 2 - Mean French language development scores with standard errors

| | Mean | SE |
|---|---|---|
| Acquisition | 1.18 years | .114 |
| Fluent | 3.6 years | .191 |
| Reading | 5.65 years | .109 |
| Fluent reading | 7.51 years | .159 |

***Training phase***

Overall accuracy in the Training phase (*MEAN* = 91.36%, *SE* = .81) suggest that participants were relatively successful in matching the target word with its translation.

**Response time.** Performance on training trials is presented in Table 4. Only correct responses were included in RT analysis. Responses on trials with a French target and Croatian labels were significantly faster than responses on trials with a Croatian target and French labels, $t(115) = 6.309$, $p < .001$, $MEAN_{diff} = -111$, $SE_{diff} = 17.6$, *Cohen's d* = -.586, $BF_{10} > 100$.

**Percentage error.** Participants responded more accurate on trials with a French target and Croatian targets relative to trials with Croatian target and French targets, $t(115) = 6.906$, $p < .001$, $MEAN_{diff} = -6.948$, $SE_{diff} = 1.01$, *Cohen's d* = -.641, $BF_{10} > 100$.

***Table 4.*** Experiment 2 - Response times and percentage errors with standard errors in the training phase

| Type of trial | Response Time | | Percentage Error | |
|---|---|---|---|---|
| | Mean | SE | Mean | SE |
| French target, Croatian labels | 1161.04 | 21.97 | 5.16 | .71 |
| Croatian target, French labels | 1271.84 | 21.41 | 12.11 | 1.14 |

***Stroop task***

**Response times.** There was a significant main effect of Congruency, $F(2,230) = 28.709$, $p < .001$, $MSE = 1911.558$, $\eta^2_p = .20$, $BF_{10} > 100$. The main effect of Language was observed, $F(1,115) = 6.060$, $p < .05$, $MSE = 2016.867$, $\eta^2_p = .05$, $BF_{10} = 1.18$. Most importantly, the interaction between Congruency and Language was significant as well, $F(2,230) = 10.229$, $p < .001$, $MSE = 2119.633$, $\eta^2_p = .082$, $BF_{10} = 589.28$.

Comparisons were conducted on each language separately. Results are visible in Figure 9. For French colour words, there was a significant stimulus conflict effect (same response – identity), $t(115) = 4.578$, $p < .001$, $MEAN_{diff} = -28.453$, $SE_{diff} = 6.22$, *Cohen's d* = -.425, $BF_{10} > 100$, and response conflict effect (different response – same response), $t(115) = 3.754$, $p < .001$, $MEAN_{diff} = -21.439$, $SE_{diff} = 5.71$, *Cohen's d* = -.349, $BF_{10} = 68.5$. The overall Stroop interference effect (different response – identity) was significant, $t(115) = 8.627$, $p < .001$, $MEAN_{diff} = -49.892$, $SE_{diff} = 5.78$, *Cohen's d* = -.801, $BF_{10} > 100$. For Croatian colour words, neither the stimulus conflict effect (same response – identity), $t(115) = 1.334$, $p > .05$, $MEAN_{diff} = -8.067$, $SE_{diff} = 6.05$, *Cohen's d* = -.124, $BF_{10} = .244$, nor the response conflict effect (different response – same response), $t(115) = .555$, $p > .05$, $MEAN_{diff} = -3.176$, $SE_{diff} = 5.72$, *Cohen's d* = -.052, $BF_{10} = .12$, were significant. The overall Stroop interference effect (different response – identity) was only marginally significant, $t(115) = 1.913$, $p = .06$, $MEAN_{diff} = -11.243$, $SE_{diff} = 5.88$, *Cohen's d* = -.178, $BF_{10} = .599$.

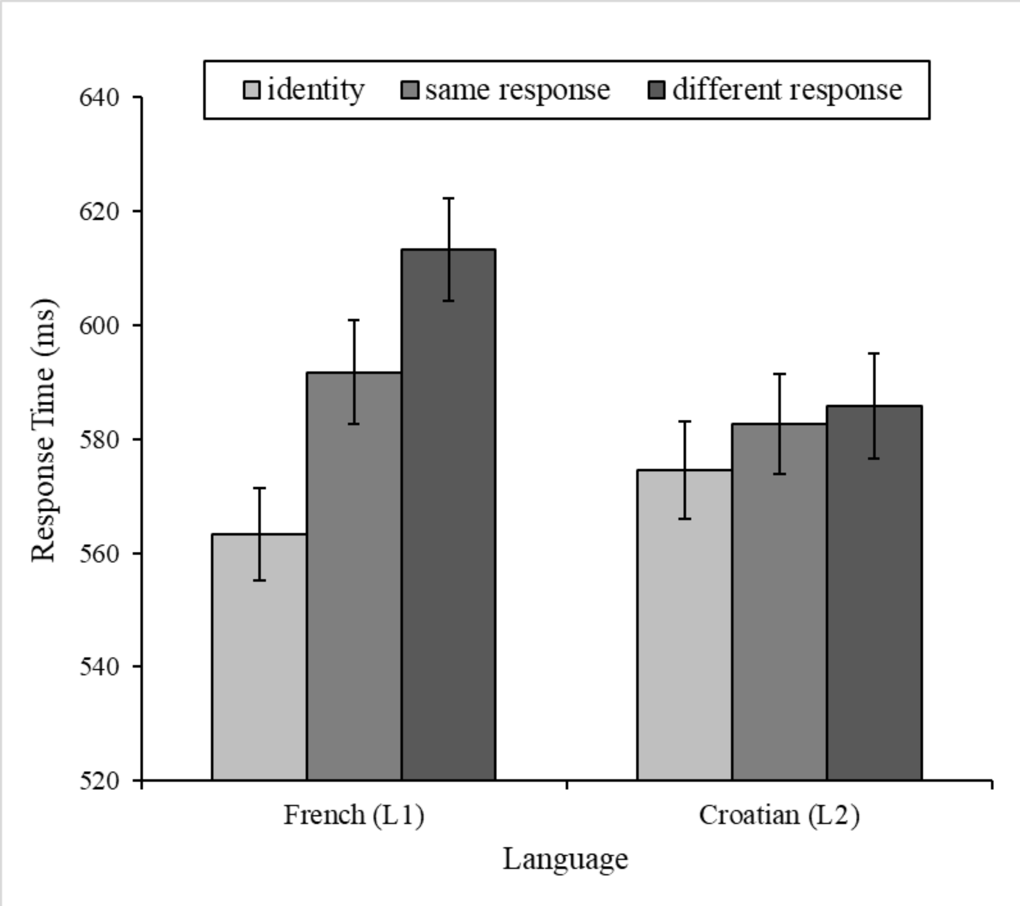


***Figure 9.*** Experiment 2 - Response times with standard errors for French and Croatian colour words in the Stroop task

The magnitudes of the stimulus conflict effect and response conflict effect were compared across languages. The magnitude of the French stimulus conflict effect (same response – identity) was significantly larger than the magnitude of the Croatian stimulus conflict effect, $t(115) = 2.24$, $p < .05$, $MEAN_{diff} = 20.4$, $SE_{diff} = 9.11$, *Cohen's d* = .208, $BF_{10} = 1.13$. The response conflict effect (different response – same response) was significantly larger for French than for Croatian colour words, $t(115) = 2.17$, $p < .05$, $MEAN_{diff} = 18.3$, $SE_{diff} = 8.41$, *Cohen's d* = .202, $BF_{10} = .987$.

**Percentage error.** There was a significant main effect of Congruency, $F(2,230) = 23.567$, $p < .001$, $MSE = 22.088$, $\eta^2 = .17$, $BF_{10} > 100$. The main effect of Language was observed, $F(1,115) = 6.358$, $p < .05$, $MSE = 22.140$, $\eta^2 = .052$, $BF_{10} = 1.56$. The interaction between Congruency and Language was significant as well, $F(2,230) = 5.646$, $p < .01$, $MSE$ = 20.567, $\eta^2 = .047$, $BF_{10} = 4.609$.

The separate analysis for French and Croatian colour words was conducted. The results are presented in Figure 10. For French colour words, no stimulus conflict effect (same response-identity) was observed, $t(115) = .081$, $p > .05$, $MEAN_{diff} = .052$, $SE_{diff} = .641$, *Cohen's d* = .007, $BF_{10} = .103$. The response conflict effect (different response-same response), $t(115) = 6.680$, $p < .001$, $MEAN_{diff} = -3.784$, $SE_{diff} = .567$, *Cohen's d* = -.62, $BF_{10} > 100$, and the overall Stroop interference effect (different response-identity), $t(115) = 6.295$, $p < .001$, $MEAN_{diff} = -3.733$, $SE_{diff} = .593$, *Cohen's d* = -.584, $BF_{10} > 100$, were significant. For Croatian colour words, stimulus conflict effect (same response-identity), $t(115) = 1.151$, $p > .05$, $MEAN_{diff} = .75$, $SE_{diff} = .651$, *Cohen's d* = .107, $BF_{10} = .196$, and the overall Stroop interference effect (different response-identity), $t(115) = 1.651$, $p > .05$, $MEAN_{diff} = -1.01$, $SE_{diff} = .611$, *Cohen's d* = -.153, $BF_{10} = .384$, failed to reach significance. The response conflict effect (different response-same response), $t(115) = 3.083$, $p < .01$, $MEAN_{diff} = -1.759$, $SE_{diff} = .570$, *Cohen's d* = -.286, $BF_{10} = 8.94$, was significant.

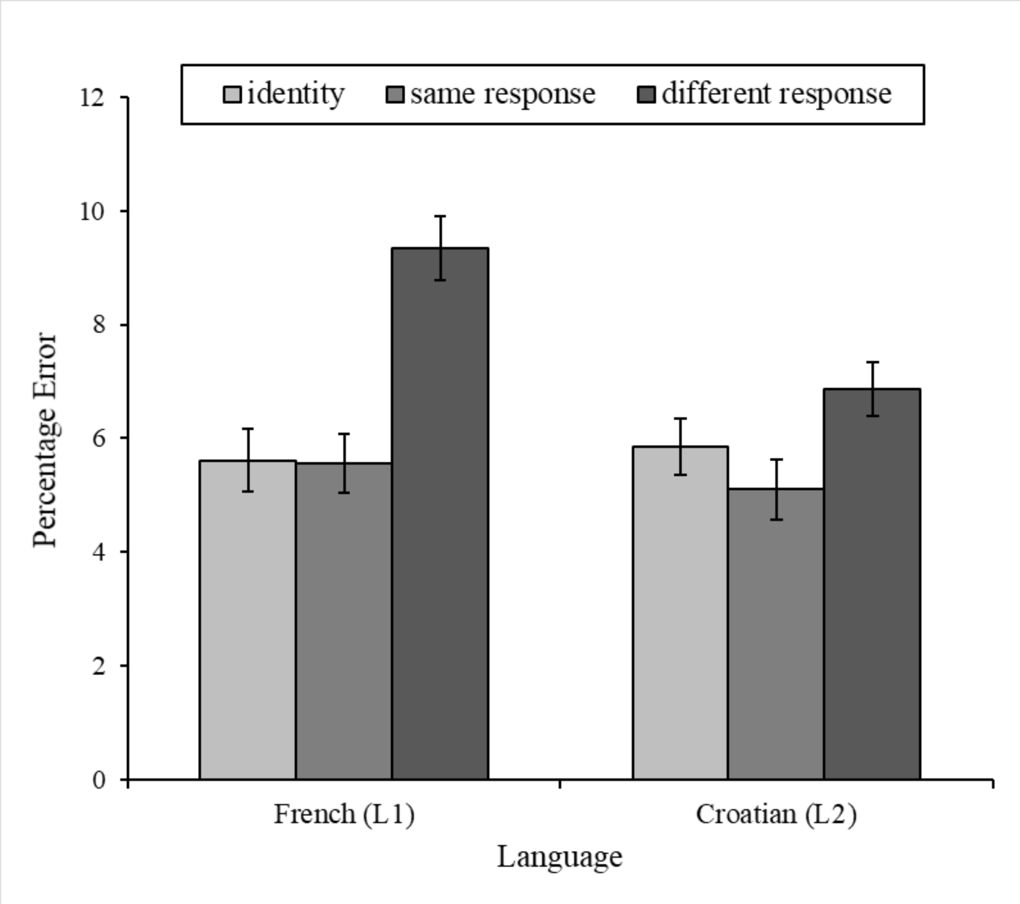


***Figure 10.*** Experiment 2 - Percentage errors with standard errors for French and Croatian colour words in the Stroop task

Again, we conducted the comparisons between the magnitudes of stimulus conflict effect and response conflict effect across languages. There was no difference in magnitude of stimulus conflict effect (same response-identity) between French and Croatian colour words, $t(115) = .773$, $p > .05$, $MEAN_{diff} = .698$, $SE_{diff} = .903$, *Cohen's d* = .072, $BF_{10} = .138$. The response conflict effect (different response-same response) was significantly larger for French than for Croatian colour words, $t(115) = 2.41$, $p < .05$, $MEAN_{diff} = 2.03$, $SE_{diff} = .842$, *Cohen's d* = .223, $BF_{10} = 1.63$.

**Discussion**

Experiment 2 aimed to investigate the source of the interference effect for novel, recently trained L2 colour words. We tested the assumptions of the Kroll & Stewart (1994) model that implies the presence of a lexical link between a novel L2 word and its L1 counterpart (i.e., presence of response conflict) exclusively in the early phases of L2 acquisition. As previously discussed, this model excludes the existence of a conceptual connection between L2 words and semantics (i.e., no stimulus conflict) for low proficient L2 speakers.

The present experiment included a training procedure that aimed to train participants with novel Croatian colour words. To establish a link between native language words and novel obscure words, we administered a task in which the word from one language (i.e., target) had to be matched with its counterpart from another language (i.e., label). This training phase was relatively short (5-10 minutes) and simple. Our L2 training procedure was substantially shorter and less complex that the one administered by Altarriba and Mathis (1997; for full description, see Introduction). Their procedure included both visual and auditory presentation of L1-L2 word pairs, as well as a series of quizzes that aimed to emphasize the semantic link between English (L1) and Spanish (L2) counterparts. For instance, participants had to write in the English word that corresponded to a given Spanish word or had to complete simple English close-ended sentences with a matching Spanish word. In contrast, we opted to use visual presentation of L1-L2 word pairs exclusively and fast-responding (up to 3 seconds) training trials. Participants responded faster and more accurately when the target was a French word, presented together with four Croatian colour word labels printed in the corresponding colour. For instance, when the target was the French word “vert” (i.e., green), a Croatian counterpart (i.e., “zelena”) was coloured in the corresponding colour (i.e., green). Therefore, when responding, it seems plausible that participants relied more on the colour of the Croatian label than on the label itself. This could facilitate responding if participants used this “colour label” cue rather than word form itself. Consequentially, Croatian word forms might not be memorized well enough, and more importantly, might not be sufficiently connected to their semantic representations. This shortcoming was addressed in Experiment 3.

The training phase was followed by the bilingual Stroop task with intermixed native French and recently learned Croatian colour words. The semantic component of the Stroop task makes it suitable for studying processes involved in second language acquisition (Altarriba & Mathis, 1997). A theoretical account of the Stroop interference paradigm concerning the automaticity hypothesis suggests that interference occurs because distracters are read automatically, but colour naming requires more attention (MacLeod, 1991; MacLeod & Dunbar, 1988; Posner & Snyder, 1975). Native language words have been integrated in the mental lexicon for a long time and have privileged access to semantics that automatically activates colour representations. In incongruent trials, both relevant (i.e., ink colour) and irrelevant (i.e., word meaning) colour representations are simultaneously activated, which produces interference. Similarly to Schmidt and colleagues (2018), we observed a significant stimulus and response conflict effect for L1 words, as an expected pattern of results.

As already discussed (see *Cognates* section in Introduction), greater interference was found for bilinguals when the colour words are similar to their counterparts in the naming language (e.g., blue/bleu). Similar words could automatically activate the corresponding word in the naming language (Dyer, 1971; Preston & Lambert, 1969). In to-be-learned novel words selection, we were attentive to use only non-cognates, with matching word length. Although there were no obvious French-Croatian colour word cognates, we chose Croatian colour words that are easily pronounced by French and have no special characters common for Croatian language (e.g., “žuta”/yellow, etc.).

Apart from similarities between L1 and to-be-learned L2, early learners possess limited L2 vocabulary knowledge. In the learning experiments, they become experts on a small set of L1-L2 translations that could be encoded on both lexical and semantic level. An important finding from early language acquisition studies suggested that novice learners can process novel words at a conceptual level even after short exposure to L2 words. This implies the presence of L2 interference effects already at early stages of language acquisition (Altarriba & Mathis, 1997). Even though certain studies already looked for and found interference effects with recently learned L2 words (Altarriba & Basnight-Brown, 2012; Altarriba & Mathis, 1997; Geukes et al., 2015), none of them tried to decompose this interference effect into underlying components. This motivated us to look closely at the source of L2 interference effect. For Croatian colour words, a marginally significant overall Stroop interference effect was found. Unfortunately, it seems that our training procedure did not result in a robust L2 congruency effect in the subsequent Stroop portion of the experiment. Even though this interference effect was not large, we decided to decompose it. Additional analyses revealed that stimulus and response conflict effects for Croatian words failed to reach statistical significance, although there was a hint toward L2 stimulus conflict.

Since our research question was to determine whether an L2 congruency effect is due to stimulus or response conflict (or both), we obviously need to produce a sufficiently larger and robust L2 congruency effect, before we can decompose it into two subcomponents (i.e., stimulus and response conflict). It seems that L2 distracters were not sufficiently processed in the Stroop task, possibly due to an insufficiently long and demanding training procedure that was supposed to ensure semantic and response processing of L2 words. However, the pattern of results hinted that there is a tendency toward an L2 stimulus conflict effect, which encouraged us to continue this line of research with adapted training procedures. In contrast, stimulus conflict is rarely observed in the errors in the 2-to-1 mapping procedure and was unsurprisingly not observed for either language in the current experiment. A response conflict effect for Croatian words was significant in errors. These findings encouraged us to continue this line of research with slightly modified experimental procedure in Experiment 3.

## Experiment 3

Experiment 3 attempted to address certain shortcomings of Experiment 2 and further test the development of the interference effect in recently trained L2. Several methodological changes were introduced into the training phase of the present experiment relative to Experiment 2, in order to increase the magnitude of L2 interference effect. The first modification concerns the structure of the training phase. The characteristics of the learning and training procedures could influence the use and activation of semantic information in the test phase (Craik & Tulving, 1975; Tulving & Thomson, 1973). For instance, Altarriba and Mathis (1997) used series of quizzes that aimed to emphasize the semantic aspect of L2 words (Altarriba & Basnight-Brown, 2012; Altarriba & Mathis, 1997). In line with this, in addition to the standard L1-L2 trials used in Experiment 2 to establish links between French and Croatian word forms, we added two new types of trial in the training phase. These two new trial types contained colour patches that had to be matched with the corresponding Croatian colour word. Therefore, by establishing a direct connection between the colour patch and corresponding L2 word, we aimed to stress the semantic aspect of recently trained Croatian words. We assumed that second language acquisition may be enhanced by emphasizing its semantic component. This modification also allowed speeding and automatizing word-colour pairings. A second modification concerns the length of training, which was slightly extended relative to Experiment 2 but remained short and did not exceed 15 minutes. Third, in the training phase of Experiment 3, colour word labels were printed in black. As previously discussed, colour word labels (e.g., “vert”) in Experiment 2 were printed in their corresponding colour (e.g., green). This manipulation had a facilitative effect for Croatian labels (e.g., “zelena” printed in corresponding green ink), since it did not require paying attention to the Croatian word. By presenting all colour word labels in black, we aimed to force participants to pay more attention to word forms, especially on recently learned Croatian colour word forms. Moreover, we tried to avoid responses based on the print colour of labels and to strengthen the lexicosemantic connections between L1 and L2 words. This new semantic link between L2 words and corresponding concepts was tested in the Stroop task that contained both novel Croatian and native French colour words. Novel words were presented either in their congruent (i.e., “learned”) ink colour (e.g., “plava” in blue) or incongruent (e.g., “plava” in green) print colour. The Stroop effect was tested immediately after learning, which rules out the possible influence of memory consolidation (Geukes et al., 2015). The final modification was the introduction of “catch” trials. As “catch” trial distracters we used two novel, random Croatian words. Participants were instructed to withhold their response when these specific distracters were presented. Thus, they were required to process the identity of the distracters on each trial within the Stroop task. This manipulation was expected to increase the Croatian congruency effect since participants were forced to attend to the distracters and distinguish the ones they had previously learned from two additional filler words (Liefooghe et al., 2020).

## Method

### *Participants*

A total of 115 University of Burgundy undergraduates (97 women, 18 men) participated in the study. The recruitment criteria and procedure were identical to that of Experiment 2. None of them spoke Croatian or a similar language and they had not studied Croatian in the school. More demographic information about the sample can be found in the Results section. Participants who already participated in Experiment 2 were not allowed to participate in Experiment 3.

### *Apparatus and material*

Identical to Experiment 2.

### *Design*

**Learning phase.** The learning phase was identical to that in Experiment 2.

**Training phase.** The training phase was conceptually similar as Experiment 2, with certain modifications. First, in addition to the two training trial types in Experiment 2, we added two additional trial types in Experiment 3. In one of them, participants were presented with a Croatian target word with four colour patches below it. In second one, a colour patch was a target that had to be matched with one of four Croatian colour words presented underneath. Second, in Experiment 3, text response labels were printed in black. Third, the number of trials was increased from 32 to 64. Each type of trial was presented 16 times, and each colour served as a target in 4 trials. All trial types used in the Training phase in Experiment 3 are shown in Figure 11.

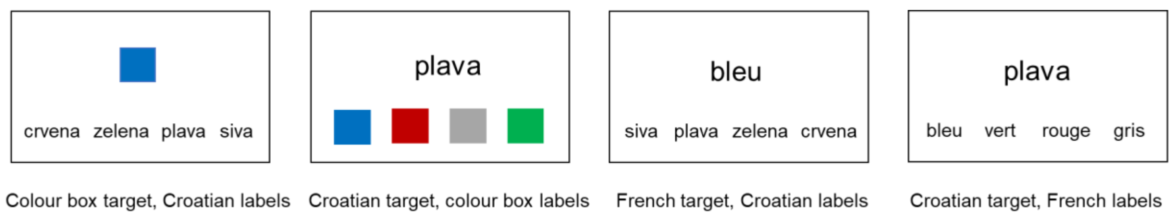


***Figure 11.*** Experiment 3 - Types of trials in the training phase

**Stroop task.** The Stroop task phase was identical to that in Experiment 2, except for one modification. In each experimental block we added 8 catch trials, that is, trials on which participants had to withhold their response. In total, there were 216 trials (including 192 experimental trials and 24 catch trials) across three experimental Stroop task blocks.

### *Procedure*

The *learning phase* was the same as the one in Experiment 2. In the *training phase*, the key assignment for response labels and timing were the same as in Experiment 2. The procedure of the *Stroop task* was identical as in Experiment 2, with the exception of the catch stimuli. As catch stimuli, we used 2 unknown Croatian words “nakon” and “mokar”. Presentation time for these stimuli was restricted to 1000 ms. If participants failed to withhold their response on these trials, the error message “Ne répondez pas à ce mot” (“Do not respond to this word”) appeared in black for 2000 ms before the next trial.

**Results**

***Language demographic***

Almost all participants indicated French as their first dominant language (98.26%) and first acquired language (95.65%). The most frequent second languages in order of dominance are English (66.96%), Spanish (20%), Arabic (4.35%), and Turkish (2.61%). Second languages in order of acquisition were English (71.3%), Spanish (6.96%), Arabic (6.09%), French (4.35%), Italian (3.48%), and Portuguese (3.48%). The estimated percentage of daily exposure to the French language was relatively high (*MEAN* = 77.05%, *SE* = 1.52). Mean French language development scores are shown in Table 5.

***Table 5.*** Experiment 3 - Mean French language development scores with standard errors

| | Mean | SE |
|---|---|---|
| Acquisition | 1.93 years | .197 |
| Fluent | 4.15 years | .20 |
| Reading | 5.77 years | .107 |
| Fluent reading | 7.68 years | .164 |

***Training phase***

Overall accuracy in the Training phase was good (*MEAN* = 89.17%, *SE* = 1.13).

**Response time.** Only correct responses were included in the RT analyses. Responses on trials with a Croatian target and colour box labels were significantly faster than responses to trials with a colour box target and Croatian labels, $t(114) = 10.58$, $p < .001$, $MEAN_{diff}$ = 154.104, $SE_{diff}$ = 14.6, *Cohen’s d* = .987, $BF_{10} > 100$, a French target with Croatian labels,

$t(114) = 10.724$, $p < .001$, $MEAN_{diff} = -176.473$, $SE_{diff} = 16.5$, *Cohen's d* = -1.00, $BF_{10} > 100$, and a Croatian target with French labels, $t(114) = 12.286$, $p < .001$, $MEAN_{diff} = -191.63$, $SE_{diff} = 15.6$, *Cohen's d* = -1.15, $BF_{10} > 100$. Responses on trials with a colour box target and Croatian labels were faster than responses on trials with a Croatian target and French labels, $t(114) = 2.194$, $p < .05$, $MEAN_{diff} = -37.526$, $SE_{diff} = 17.1$, *Cohen's d* = -.205, $BF_{10} = 1.03$.

**Percentage error.** Responses were less accurate on trials with a Croatian target and French labels than trials with a colour box target and Croatian labels, $t(114) = 2.736$, $p = .01$, $MEAN_{diff} = -2.826$, $SE_{diff} = 1.03$, *Cohen's d* = -.255, $BF_{10} = 3.58$, a Croatian target with colour box labels, $t(114) = 2.537$, $p = .01$, $MEAN_{diff} = -2.374$, $SE_{diff} = .936$, *Cohen's d* = -.237, $BF_{10} = 2.21$, and a French target with Croatian labels, $t(114) = 3.465$, $p = .001$, $MEAN_{diff} = 3.165$, $SE_{diff} = .913$, *Cohen's d* = .323, $BF_{10} = 27.4$. Response times and percentage errors for all trial types are presented in Table 6.

***Table 6.*** Experiment 3 - Response times and percentage errors with standard errors in the training phase

| Type of trial | Response Time | | Percentage Error | |
|---|---|---|---|---|
| | Mean | SE | Mean | SE |
| Colour box target, Croatian labels | 1548.45 | 25.15 | 10.10 | 1.35 |
| Croatian target, colour box labels | 1394.34 | 24.04 | 10.55 | 1.27 |
| French target, Croatian labels | 1570.82 | 23.74 | 9.76 | 1.21 |
| Croatian target, French labels | 1585.97 | 23.44 | 12.92 | 1.26 |

***Stroop task***

The performance on catch trials was analysed separately from experimental trials within the Stroop tasks. There were 8 catch trials per Stroop block (in total, 24 over the entire experiment). Participants were relatively successful in withholding their responses on catch trials (*MEAN* = 79.05%, *SE* = 1.81).

**Response times.** A significant main effect of Congruency, $F(2,228) = 27.06$, $p < .001$, $MSE = 2446.10$, $\eta_p^2 = .19$, $BF_{10} > 100$, was found. There was no main effect of Language, $F(1,114) = 1.87$, $p > .05$, $MSE = 3249.35$, $\eta_p^2 = .02$, $BF_{10} = .265$. The interaction between Congruency and Language was significant, $F(2,228) = 14.08$, $p < .001$, $MSE = 3018.30$, $\eta_p^2 = .11$, $BF_{10} > 100$.

The comparisons between different congruency trials were conducted within languages. The results are shown at Figure 12. For French colour words, there was a significant stimulus conflict (same response-identity) effect, $t(114) = 3.60$, $p < .001$, $MEAN_{diff} = -26.919$, $SE_{diff} = 7.48$, *Cohen's d* = .33, $BF_{10} = 41.32$, and response conflict (different response-same response) effect, $t(114) = 5.067$, $p < .001$, $MEAN_{diff} = -33.716$, $SE_{diff} = 6.65$, *Cohen's d* = .47, $BF_{10} > 100$. The overall Stroop interference (different response-identity) effect was also significant, $t(114) = 8.648$, $p < .001$, $MEAN_{diff} = -60.635$, $SE_{diff} = 7.01$, *Cohen's d* = .81, $BF_{10} > 100$. For Croatian colour words, neither a stimulus conflict (same response-identity) effect, $t(114) = 1.195$, $p > .05$, $MEAN_{diff} = -8.858$, $SE_{diff} = 7.41$, *Cohen's d* = .11, $BF_{10} = .207$, nor a response conflict (different response-same response) effect, $t(114) = .264$, $p > .05$, $MEAN_{diff} = 1.667$, $SE_{diff} = 6.30$, *Cohen's d* = .02, $BF_{10} = .107$, reached significance. The overall Stroop interference (different response-identity) effect was not significant, $t(114) = 1.123$, $p > .05$, $MEAN_{diff} = -7.191$, $SE_{diff} = 6.40$, *Cohen's d* = .10, $BF_{10} = .191$.

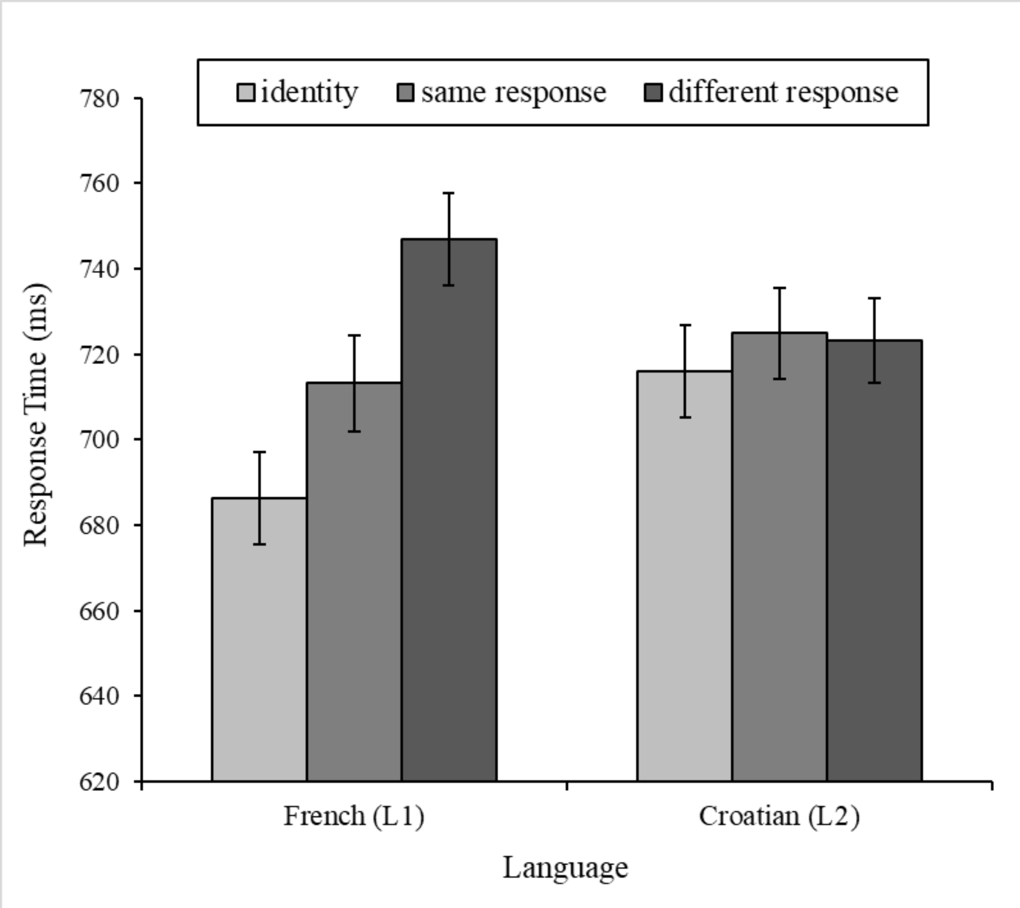


***Figure 12.*** Experiment 3 - Response times with standard errors for French and Croatian colour words in the Stroop task

The comparison across languages revealed that there was no evidence for significant difference in the magnitude of stimulus conflict (same response-identity) effect between French and Croatian words, $t(114) = 1.661$, $p > .05$, $MEAN_{diff} = 18.061$, $SE_{diff} = 10.9$, *Cohen's d* = .155, $BF_{10} = .392$. The response conflict (different response-same response) effect, t(114) = 3.482, $p = .001$, $MEAN_{diff} = 35.382$, $SE_{diff} = 10.2$, *Cohen's d* = .325, $BF_{10} = 28.9$, and the overall Stroop interference effect, $t(114) = 5.529$, $p < .001$, $MEAN_{diff} = 53.443$, $SE_{diff} = 9.67$, *Cohen's d* = .516, $BF_{10} > 100$, were more pronounced for French than for Croatian words.

**Percentage error.** The main effect of Congruency was observed, $F(2,228) = 33.776$, $p < .001$, $MSE = 41,22$, $\eta_p^2 = .23$, $BF_{10} > 100$. There was neither significant main effect of Language, $F(1,114) = .64$, $p > .05$, $MSE = 33.51$, $\eta_p^2 = .02$, $BF_{10} = .119$, nor a significant

Congruency * Language interaction, $F(2,228) = 2.51$, $p > .05$, $MSE = 29.11$, $\eta_p^2 = .02$, $BF_{10} = .215$.

Despite the lack of interaction, we conducted comparisons within languages. The results are shown at Figure 13. For French colour words, there was no significant stimulus conflict (same response-identity) effect, $t(114) = .951$, $p > .05$, $MEAN_{diff} = .669$, $SE_{diff} = .70$, *Cohen's d* = .09, $BF_{10} = .161$. The response conflict (different response-same response) effect, $t(114) = 6.198$, $p < .001$, $MEAN_{diff} = -5.539$, $SE_{diff} = .89$, *Cohen's d* = .58, $BF_{10} > 100$, and the overall Stroop interference (different response-identity) effect, $t(114) = 5.686$, $p < .001$, $MEAN_{diff} = -4.869$, $SE_{diff} = .86$, *Cohen's d* = .53, $BF_{10} > 100$, were significant. For Croatian colour words, the stimulus conflict (same response-identity) effect was not significant, $t(114) = .568$, $p > .05$, $MEAN_{diff} = .365$, $SE_{diff} = .64$, *Cohen's d* = .05, $BF_{10} = .132$. The response conflict (different response-same response) effect, $t(114) = 4.366$, $p < .001$, $MEAN_{diff} = -3.452$, $SE_{diff} = .79$, *Cohen's d* = .41, $BF_{10} > 100$, and the overall Stroop interference (different response-identity) effect, $t(114) = 3.975$, $p < .001$, $MEAN_{diff} = -3.087$, $SE_{diff} = .78$, *Cohen's d* = .37, $BF_{10} > 100$, were significant.

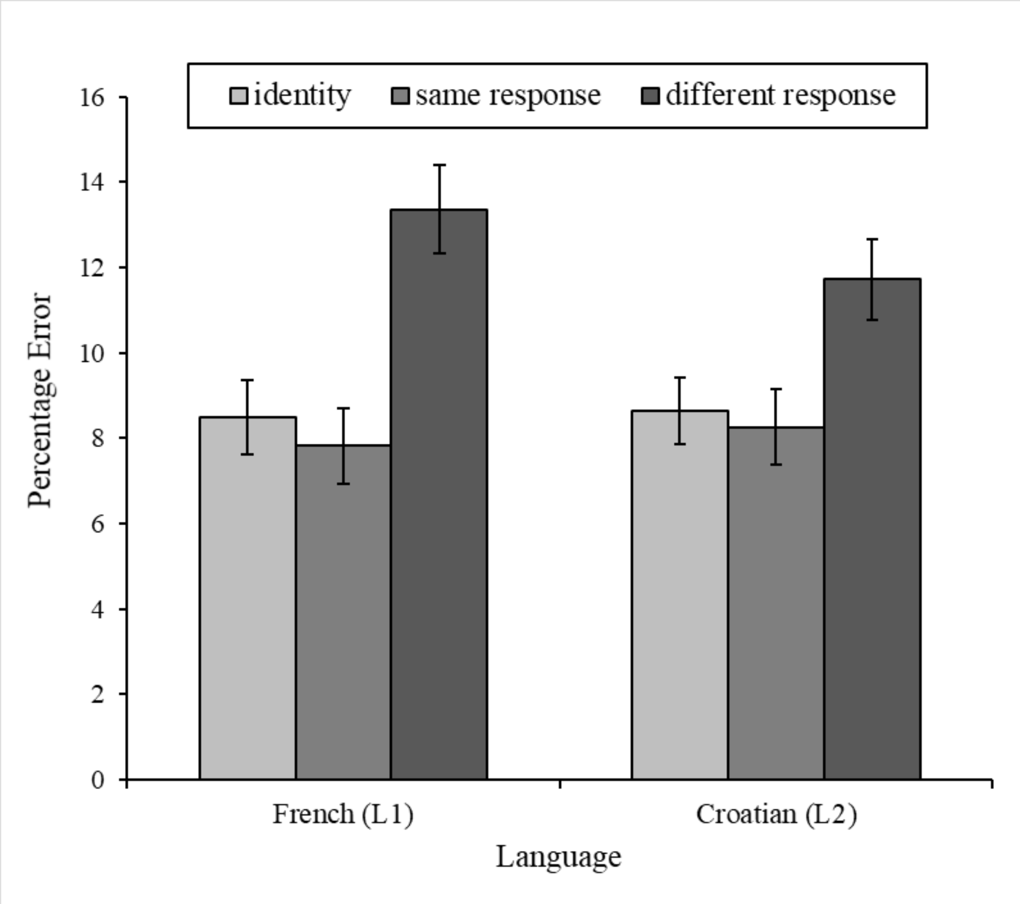


***Figure 13.*** Experiment 3 - Percentage errors with standard errors for French and Croatian colour words in the Stroop task

Comparison across languages revealed no difference in the magnitude of the stimulus conflict (same response-identity) effect between French and Croatian colour words, $t(114) = .332$, $p > .05$, $MEAN_{diff} = -.304$, $SE_{diff} = .917$, *Cohen's d* = -.031, $BF_{10} = .109$. There was a marginally significant difference in the magnitude of the response conflict (different response-same response) effect between languages, $t(114) = 1.993$, $p = .05$, $MEAN_{diff} = 2.087$, $SE_{diff} = 1.05$, *Cohen's d* = .186, $BF_{10} = .696$. No significant difference in overall Stroop interference was observed between French and Croatian colour words, $t(114) = 1.699$, $p > .05$, $MEAN_{diff} = 1.782$, $SE_{diff} = 1.05$, *Cohen's d* = .158, $BF_{10} = .417$.

## Discussion

Several methodological shortcomings of Experiment 2 were addressed in the present Experiment 3 (see Introduction). One of them concerns the structure and length of the training procedure. As already discussed, in the training procedure we introduced two new trial types that aimed to enhance the connections between novel Croatian word forms and their corresponding semantic representations. For instance, a novel type of trial that contained a Croatian target and colour patch labels was responded to significantly faster than other types of trial. This might suggest that colour patch labels facilitated activation of underlying semantic representation when a novel L2 word is presented. Apart from separate training trials analysis, the overall accuracy in the training phase was satisfying, which suggests that participants were successful in matching novel L2 words with their L1 translations or corresponding semantic representations. Furthermore, the length of the training phase in the present experiment was twice as long as in Experiment 2 but remained relatively short (around 10-12 minutes). This increase in the length of the training procedure was reasonable considering the insertion of two additional types of training trials. Also, colour word labels were printed in black, which forced participants to pay more attention to word forms, rather than only relying on ink colour. This was of crucial importance since the retrieval of novel L2 word forms was expected to influence the performance in the Stroop task. Finally, in the Stroop task we inserted "catch" trials (i.e., novel random Croatian words) that were supposed to boost the L2 interference effect.

As in previous experiments, performance in the Stroop portion was analysed separately for native French (L1) and recently trained Croatian (L2) colour words. As expected, a strong overall interference effect was produced for L1 words. After decomposing it, the stimulus conflict effect and response conflict effect were significant for French colour words, which is consistent with previous findings (e.g., Experiment 2, see also Schmidt et al., 2018). However, of particular interest was the interference effect for novel L2 (Croatian) words that received a considerable amount of training. Despite all aforementioned modifications introduced in the training procedure, no substantial Stroop effect was obtained for Croatian colour words. Surprisingly, our adapted training procedure with the addition of "catch" trials in the Stroop task did not produce robust L2 interference. Therefore, our research question that concerns the source of this effect could not be addressed. Related to this, neither stimulus conflict nor response conflict effects were observed for novel L2 words in the response times. However, similarly as in Experiment 2, a significant L2 response conflict was observed in the errors. This could possibly imply that a response conflict effect (at least for errors) occurs in very early phases of L2 acquisition.

Taken together, the results of Experiment 3 suggest that the novel L2 words do not demonstrate sufficiently strong connections with their L1 lexical representation (i.e., no L2 response conflict effect). If so, L2 words would be automatically translated into their corresponding L1 translations that would in turn lead to the activation of the response associated with that colour. Another observation concerns the absence of an L2 stimulus conflict effect, suggesting that recently trained L2 words do not directly access a semantic representation of the colour word they are associated with. One explanation of the observed pattern concerns the features of L2 word training. Since our training manipulation did not produce sufficiently large L2 interference effect, it is plausible that the word training portion was either too short or too complex. In other words, participants were not trained enough due to the small number of training trials, or complexity of training trials themselves, which contained multiple response alternatives. These issues were addressed in Experiment 4.

## Experiment 4

Experiment 4 continued the present line of research that aims to investigate the presence the L2 interference effect, with a particular focus on its subcomponents (i.e., stimulus and response conflict). In Experiment 4, we tested if a substantial L2 interference effect is more likely to occur with an increased length of L2 training. From a methodological point of view, Experiment 4 introduced several modifications relative to Experiment 3. First, the length of the training procedure was extended from 64 (Experiment 3) to 576 (Experiment 4) training trials. By drastically increasing the length and amount of L2 training trials, we aimed to strengthen the connection between novel L2 words and their underlying semantic representations. Second, each training trial contained two response labels (Experiment 4) instead of four (Experiment 2 and 3). We expected that this decrease in the number of response alternatives could accelerate responding and automatize the formation of links between L2 words and semantics. In combination with longer training, this manipulation was supposed to lead to more robust L2 interference effects relative to previous experiments.

To sum up, Experiment 4 was concerned with two questions: 1) will a L2 congruency effect emerge as a result of more extensive L2 word training, and 2) will the increase in length of the training, along with reduced number of response alternatives per trial, strengthen semantic links for L2 words? If this is true, our improved training procedure should produce a substantial L2 congruency effect, which allows for decomposing it. After conducting separate analyses on different congruency trials (i.e., identity, same response, different response trials), we expected to observe a significant difference in response latencies between identity and same response trials. In other words, if our manipulation is efficient, we should be able to observe an L2 stimulus conflict, which results from connections between novel L2 words and their corresponding semantic representations.

## Method

### *Participants*

A total of 122 participants (85 women and 37 men) were recruited online via the prolific.ac website. An additional six submissions were rejected by the experimenter due to technical issues or inappropriate completion time (too short or too long). A further 62 submissions were incomplete, and therefore excluded. Over half of these dropouts (34) did not complete the survey portion of the task. We do not know how far the remaining dropouts (28) progressed in the task, as the Psytoolkit server does not store incomplete data (e.g., it is possible that a technical error occurred when launching the experimental portion of the task). We note that dropout rates like this are typical for online data collection, but results are generally very similar to lab-collected data (Crump et al., 2013). The selection was based on following criteria: speaking English as a native language, having normal or corrected-to-normal colour vision, and having no prior knowledge of the Croatian language. Only 27.05% of participants were students. Most of them were born in United Kingdom (67.21%), United States (12.3%), Canada (7.38%), and Ireland (4.1%). Similarly, country of residence for a majority of participants was United Kingdom (68.03%), United States (13.11%), Canada (9.02%), and Ireland (4.1%). More demographic information about the sample can be found in the Results section. Participants were paid £3.5 for participation in the experiment, which lasted approximately 30 minutes.

***Apparatus and materials***

The experiment was programmed in the Psytoolkit (Stoet, 2010, 2017) programming language and designed to work on a PC. As in previous experiments, prior to the experimental portion of the experiment, participants completed a short survey concerning their language background. The structure of the survey was the same as in Experiment 2 and Experiment 3, but it was written in English. The purpose of this survey was to assure that all participants had the target language dominance (i.e., they are native English speakers) and no previous experience with the Croatian language.

***Design***

**Learning phase.** The learning phase was identical as in previous experiments, except that in Experiment 4 participants were presented with the Croatian-English colour word pairs ("crvena"-"red", "plava"-"blue", "zelena"-"green", and "siva"-"grey"). The length of the learning phase and the instructions remained unchanged.

**Training phase.** The training phase started immediately after the learning phase. There were four types of trials in the Training phase: 1) Croatian target with English labels, 2) Croatian target with colour box labels, 3) English target with Croatian labels, and 4) Colour box target with Croatian labels. These trials are illustrated in Figure 14. There were 144 trials for each trial type, for a total of 576 training trials. As mentioned, compared to Experiment 2 and Experiment 3, each training trial in Experiment 4 contained two possible response labels. Participants responded by pressing the F-key for the left response option and the J-key for the right response option. They were instructed to respond as quickly and accurately as possible.


***Figure 14.*** Experiment 4 - Types of trials in the training phase

**Stroop task.** The Stroop task was identical in all aspects to Experiment 3, with a few minor exceptions. First, the L1 colour words were English (i.e., "red", "blue", "green", "grey"). Second, the number of catch trials per Stroop block varied (i.e., it was not limited to 8 as in Experiment 3).

***Procedure***

The procedure was identical to that in Experiment 3 with the following exceptions. Except for English L1 stimuli rather than French, the *learning phase* was identical as Experiment 3. The *training phase* was similar to the prior experiments. On each trial, there was a fixation point displayed in the centre of the screen for 500 ms. Next, the target appeared in the centre of the screen together with two labels representing two response options located below the target on the left (-200, 200) and on the right (200, 200). The locations of the response options were randomized from one trial to the next. The target and the labels remained on the screen for 3000 ms or until a response was registered. The next trial began immediately following a correct response. If the participant made an error or failed to respond in 3000 ms, the message “Incorrect” or “Too slow”, respectively, appeared in black for 1000 ms before the next trial. The procedure for experimental *Stroop* trials was identical to that in Experiment 3.

## Results

### *Language demographics*

Language demographic data based on the survey responses were analysed. Almost all participants indicated English as their first dominant language (95.08%), and as their first language in order of acquisition (98.36%). The most frequent second language in order of dominance was French (30.14%), followed by German (17.81%) and Spanish (16.44%). Only 73 participants (59.84%) provided information on their second dominant language. Information on second language in order of acquisition were obtained from 69 participants (56.56%). In order of acquisition, French was rated as a second languages by 40.58% of participants, followed by Spanish (14.49%). Other languages, such as Japanese, Hindi, Punjabi, Turkish, Italian, Irish, and Welsh were also noted as second languages both in order of dominance and acquisition in small percentages. Participants are highly exposed to English language; 95.90% of them estimated this exposure rate on daily basis as 81-100%, and 4.10% of them as 61-80%. Mean English language scores are presented in Table 7. None of the participants studied Croatian in school. All of them self-rated their Croatian knowledge as 1 (“not at all”) and were not able to translate the given Croatian words.

***Table 7.*** Experiment 4 - Mean English development scores with standard errors

| | Mean | | SE |
|---|---|---|---|
| Acquisition | .73 | Years | .077 |
| Fluent | 3.83 | Years | .197 |
| Reading | 4.11 | Years | .096 |
| Fluent reading | 6.22 | Years | .168 |

***Training phase***

Overall accuracy in the Training phase was high (*MEAN* = 95.75%, *SE* = .286).

**Response time.** Only correct trials were included into RT analysis. Participants responded significantly faster on trials with a Croatian target and colour box labels than to other types of trials: Croatian target with English labels, $t(121) = 24.37$, $p < .001$, *MEANdiff* = 149.05, $SE_{diff} = 6.12$, *Cohen's d* = 2.21, $BF_{10} > 100$, English target with Croatian labels, $t(121) = 17.23$, $p < .001$, *MEANdiff* = -110.73, $SE_{diff} = 6.43$, *Cohen's d* = -1.56 $BF_{10} > 100$ and Colour box target with Croatian labels, $t(121) = 17.27$, $p < .001$, $MEAN_{diff} = -93.10$, $SE_{diff} = 5.39$, *Cohen's d* = -1.56, $BF_{10} > 100$. Responses on trials with a Croatian target and English labels were slower than responses on trials with an English target and Croatian labels, $t(121) = 6.09$, $p < .001$, *MEANdiff* = 38.32, $SE_{diff} = 6.29$, *Cohen's d* = .552, $BF_{10} > 100$, and Colour box targets with Croatian labels, $t(121) = 7.60$, $p < .001$, *MEANdiff* = 55.94, $SE_{diff} = 7.36$, *Cohen's d* = .688, $BF_{10} > 100$. Responses were significantly slower for trials with an English target and Croatian labels than trials with a Colour box target and Croatian labels, $t(121) = 2.73$, $p = .01$, *MEANdiff* = 17.63, $SE_{diff} = 6.46$, *Cohen's d* = .247, $BF_{10} = 3.45$.

**Percentage error.** Responses were significantly less accurate on trials with a Croatian target and English labels, relative to trials with a Croatian target and colour box labels, $t(121) = 8.44$, $p < .001$, *MEANdiff* = 2.04, $SE_{diff} = .242$, *Cohen's d* = .764, $BF_{10} > 100$, English target with Croatian labels, $t(121) = 9.54$, $p < .001$, *MEANdiff* = 2.72, $SE_{diff} = .285$, *Cohen's d* = .864, $BF_{10} > 100$, and Colour box target with Croatian labels, $t(121) = 8.86$, $p < .001$, *MEANdiff* = 2.36, $SE_{diff}$ = .267, *Cohen's d* = .802, $BF_{10}$ > 100. Responses on trials with a Croatian target and colour box labels were less accurate relative to trials with an English target and Croatian labels, *t*(121) = 2.95, *p* < .01, *MEANdiff* = .68, $SE_{diff}$ = .232, *Cohen's d* = .267, $BF_{10}$ = 6.096. Response times and percentage errors for all trial types are presented in Table 8.

***Table 8.*** Experiment 4 - Response times and percentage errors with standard errors in the training phase

| Type of trial | Response Time | | Percentage Error | |
|---|---|---|---|---|
| | Mean | SE | Mean | SE |
| Croatian target, English labels | 1118.19 | 13.25 | 6.03 | .38 |
| Croatian target, colour box labels | 969.14 | 12.47 | 3.99 | .30 |
| English target, Croatian labels | 1079.87 | 12.99 | 3.3 | .29 |
| Colour box target, Croatian labels | 1062.24 | 13.39 | 3.67 | .31 |

***Stroop task***

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

**Response time.** For response times, only correct responses were analysed, with no other trims. Data are presented in the Figure 15. We conducted a congruency (identity vs. same response vs. different response) by language (English vs. Croatian) within-subject repeated measures ANOVA. The main effect of Congruency was significant; *F*(2, 242) = 114.83, *p* < .001, *MSE* = 2826.67, $\eta_p^2$ = .49, $BF_{10}$ > 100. The main effect of Language was marginally significant; *F*(1,121) = 3.77, *p* = .05, *MSE* = 2906.55, $\eta_p^2$ = .03, $BF_{10}$ = .297. A significant Congruency * Language interaction was observed, *F*(2, 242) = 13.16, *p* < .001, *MSE* = 3103.40, $\eta_p^2$ = .10, $BF_{10}$ > 100.

Considering the significant interaction, we compared the different types of trials separately on each language. For English colour words, we observed a significant stimulus conflict effect (same response – identity); $t(121) = 7.33$, $p < .001$, $MEAN_{diff} = -53.5$, $SE_{diff} = 7.30$, *Cohen's d* = -.66, $BF_{10} > 100$, and response conflict effect (different response – same response); $t(121) = 6.39$, $p < .001$, $MEAN_{diff} = -44.4$, $SE_{diff} = 6.94$, *Cohen's d* = -.579, $BF_{10} > 100$. For Croatian colour words, stimulus conflict was significant; $t(121) = 4.96$, $p < .001$, $MEAN_{diff} = -35.2$, $SE_{diff} = 7.09$, *Cohen's d* = -.449, $BF_{10} > 100$, while response conflict showed a tendency toward statistical significance, $t(121) = 1.76$, $p = .08$, $MEAN_{diff} = -11.6$, $SE_{diff} = 6.60$, *Cohen's d* = -.159, $BF_{10} = .449$. The Stroop interference effect (different response – identity) was significant for both English, $t(121) = 15.05$, $p < .001$, $MEAN_{diff} = -97.9$, $SE_{diff} = 6.50$, *Cohen's d* = -1.36, $BF_{10} > 100$, and Croatian colour words, $t(121) = 6.37$, $p < .001$, $MEAN_{diff} = -46.8$, $SE_{diff} = 7.35$, *Cohen's d* = -.577, $BF_{10} > 100$.

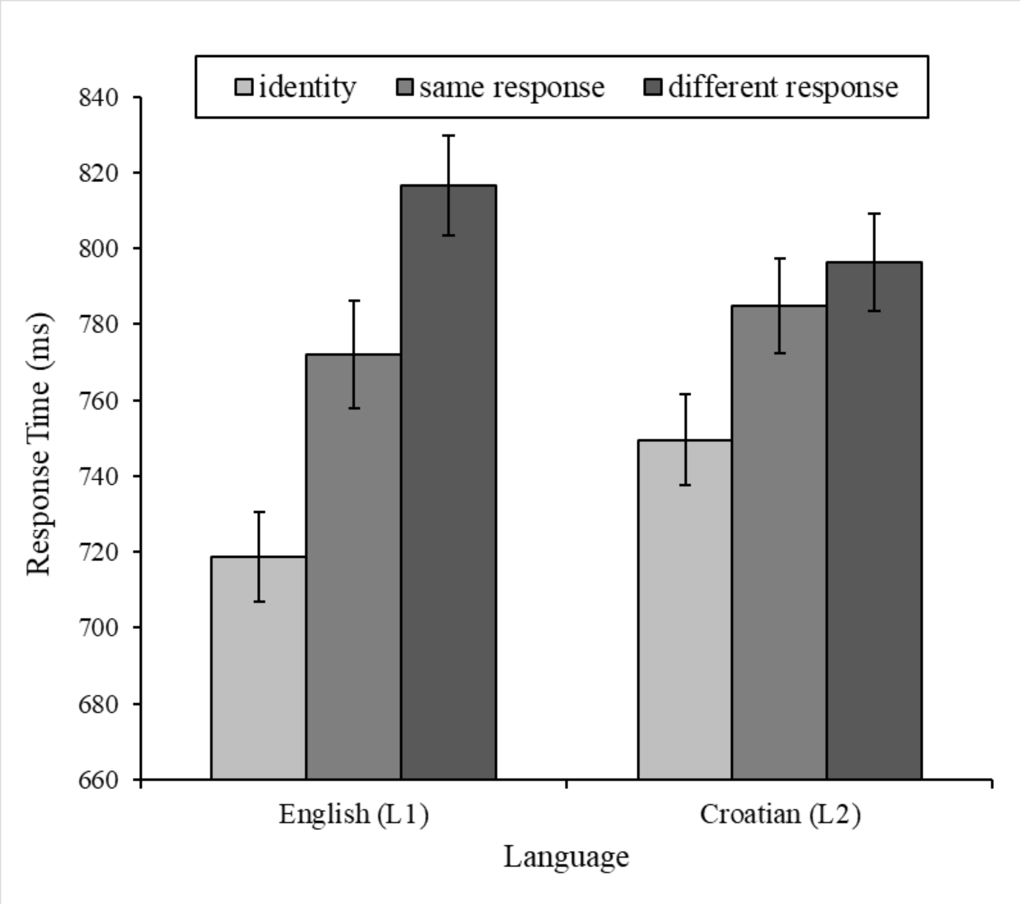


***Figure 15.*** Experiment 4 - Response times with standard errors for English and Croatian colour words in the Stroop task

There was no difference in the magnitude of stimulus conflict between English and Croatian words, $t(121) = 1.80$, $p > .05$, $MEAN_{diff} = 18.3$, $SE_{diff} = 10.2$, *Cohen's d* = .163, $BF_{10}$ = .48. However, the response conflict effect was larger for English words than for Croatian words, $t(121) = 3.19$, $p < .01$, $MEAN_{diff} = 32.8$, $SE_{diff} = 10.3$, *Cohen's d* = .289, $BF_{10} = 11.87$. The overall Stroop interference effect was larger for English than for Croatian colour words, $t(121) = 5.209$, $p < .001$, $MEAN_{diff} = 51.07$, $SE_{diff} = 9.8$, *Cohen's d* = .472, $BF_{10} > 100$.

**Percentage error.** For error percentages, trials in which participants failed to respond before the deadline were excluded (1.6% of trials). Percentage error data are presented in Figure 16. Again, a congruency (identity vs. same response vs. different response) by language (English vs. Croatian) within-subjects repeated measures ANOVA was conducted. The main effect of Congruency was observed, $F(2, 242) = 31.84$, $p < .001$, $MSE = 38.34$, $\eta_p^2$

= .21, $BF_{10} > 100$. However, there was no main effect of language, $F(1,121) = 2.72$, $p > .05$, $MSE = 26.39$, $\eta_p^2 = .02$, $BF_{10} = .203$. The interaction between congruency and language was significant, $F(2, 242) = 4.63$, $p = .01$, $MSE = 32.00$, $\eta_p^2 = .04$, $BF_{10} = 1.80$.

We again compared performance on the different types of trials across languages. For English colour words, stimulus conflict was not significant, $t(121) = .31$, $p > .05$, $MEAN_{diff} = -.224$, $SE_{diff} = .727$, *Cohen's d* = -.028, $BF_{10} = .105$, but response conflict was significant, $t(121) = 7.30$, $p < .001$, $MEAN_{diff} = -5.11$, $SE_{diff} = .70$, *Cohen's d* = -.661, $BF_{10} > 100$. Similarly for Croatian colour words, stimulus conflict was not significant, $t(121) = .09$, $p > .05$, $MEAN_{diff} = -.069$, $SE_{diff} = .804$, *Cohen's d* = -.008, $BF_{10} = .101$, while response conflict was significant, $t(121) = 3.41$, $p = .001$, $MEAN_{diff} = -2.49$, $SE_{diff} = .729$, *Cohen's d* = -.309, $BF_{10} = 23.22$. The overall Stroop interference was significant for both English, $t(121) = 7.34$, $p < .001$, $MEAN_{diff} = -5.33$, $SE_{diff} = .727$, *Cohen's d* = -.664, $BF_{10} > 100$, and Croatian colour words, $t(121) = 2.98$, $p < .01$, $MEAN_{diff} = -2.56$, $SE_{diff} = .857$, *Cohen's d* = -.27, $BF_{10} = 6.73$.

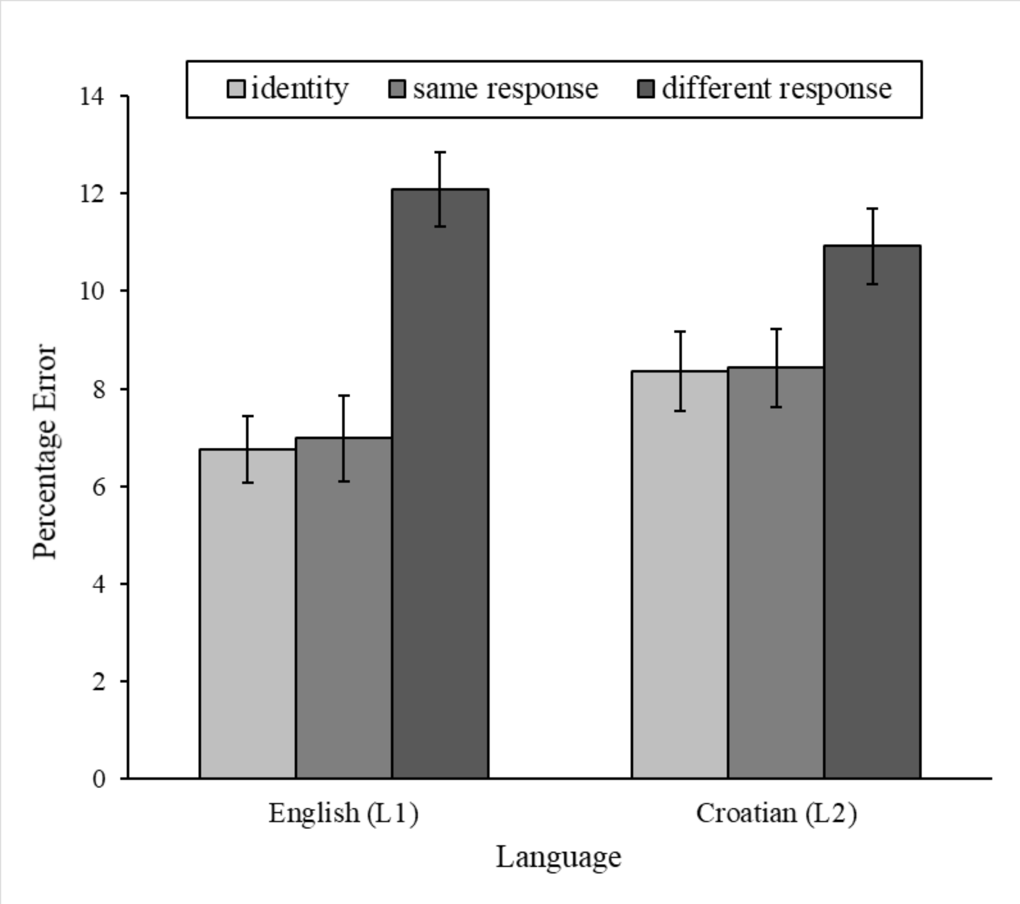


***Figure 16.*** Experiment 4 - Percentage errors with standard errors for English and Croatian colour words in the Stroop task

No significant difference between English and Croatian stimulus conflict was observed, $t(121) = .16$, $p > .05$, $MEAN_{diff} = .155$, $SE_{diff} = .974$, *Cohen's d* = .014, $BF_{10} = .102$. However, the response conflict effect, $t(121) = 2.49$, $p = .01$, $MEAN_{diff} = 2.62$, $SE_{diff} = 1.05$, *Cohen's d* = .226, $BF_{10} = 1.95$, and the overall Stroop interference effect, $t(121) = 2.649$, $p = .01$, $MEAN_{diff} = 2.773$, $SE_{diff} = 1.05$, *Cohen's d* = .24, $BF_{10} = 2.83$, were significantly larger for English words than for Croatian words.

**Discussion**

Experiment 4 follows the line of research that investigated the congruency effect in novel, recently trained L2. In order to establish a link between Croatian words and their L1 counterparts, we used learning and training procedures identical as in Experiment 2 and Experiment 3. Since previous studies did not show an observable L2 interference effects, we decided to extend and simplify the training procedure. First, the increase in the number of training trials relative to the previous experiment (from 64 to 576) was noticeable and it resulted in an increase of its duration (around 20 minutes). Second, instead of four response options, Experiment 4 used only two possible responses. Response latencies are influenced by the number of response alternatives among which participants have to decide. Trials with two response alternatives create two possible target-response associations (one target-response association which is correct and another target-response association which is incorrect). In contrast, if the task has four response alternatives, it activates four possible target-response associations. Only one of them is correct, while another three target-response associations have to be discarded. Selection among larger number of response alternatives slows down the selection process (Hick, 1952; Smith, 1982). Following this rationale, our task with two alternatives accelerated response processing and automatized the formation of links between L2 and its either lexical (L1) or semantic (colour) counterpart.

These modifications led to an interesting insight on the source of interference in novel languages. Again, significant stimulus and response conflict effects were observed for native English colour words. Contrary to Experiments 2 and 3, recently trained Croatian colour words produced a significant Stroop interference effect, which allowed us to (more meaningfully) decompose this congruency effect into stimulus and response conflict. A significant stimulus conflict effect and marginally significant response conflict effect in response times were evidenced. The stimulus conflict effect in response times observed for Croatian colour words could be interpreted as a result of early, semantic processes (Glaser & Glaser, 1989; MacKinnon et al., 1985) in word acquisition. It seems plausible that two dimensions of novel words (i.e., ink colour and word meaning) have been associatively connected over the learning and training phase. This association speeds up the responses on congruent trials in which the colour and word match. In contrast, on incongruent trials, the concurrent activation of the colour and word interfere at a semantic level and produce a difference in response latencies between identity (i.e., when colour and word match and they are assigned to the same key) and same response (i.e., when colour and word mismatch but they are assigned to the same key) trials. The response conflict effect for Croatian colour words was not significant in responses times (although trending in the expected direction) but was significant in the errors. These results are incompatible with models that assume exclusive semantic mediation early in language learning, as response conflict was observed in L2, at least for errors. The assumptions of the Kroll and Stewart (1994) model are only partially confirmed. According to this model, the novel L2 words are linked to their L1 counterparts exclusively (therefore assuming the presence of the response conflict only), at least at early stages of L2 acquisition. The L2 response conflict in the errors confirms this notion. Furthermore, the model assumes that the formation of a semantic link (i.e., the presence of the stimulus/semantic conflict) between an L2 word and its corresponding semantic representation occurs eventually, with sufficiently high language proficiency, but not so early on. Our results hint at the reverse pattern: a stimulus conflict effect was observed in L2 for response times, with faster responses on identity trials relative to same response trials. The responses were facilitated on standard congruent trials (i.e., identity trials) in L2, suggesting the presence of relatively strong link between novel word form (e.g., “plava”) and its underlying semantic representation (e.g., colour blue), similarly as in L1. The Croatian colour words used in the experiment were likely to be considered as semantic associates, and not only as translations of L1 words (Liefooghe et al., 2020). Therefore, the presence of stimulus conflict effect for novel Croatian words could be due to their direct relation to the conceptual representation of a particular colour, formed immediately after short training.

## Experiment 5

A similar pattern of results obtained for L1 and L2 colour words (see Experiment 4) suggested that semantic representation could be accessed by novel words trained with their native language equivalents. Our further investigations aimed to test whether colour words and their colour associates rely on the same processing route that could facilitate acquisition of L2 words.

The present experiment continues this line of research on L2 acquisition and investigating the source of interference in novel L2 words. However, instead of colour words (see Experiment 1-4), we chose colour associates as to-be-learned L2 words, as well as their L1 translations. As already discussed, Klein (1964) measured the interference produced by different types of stimuli in a Stroop colour identification task. One of the word types were colour associates, that is, words inherently associated with colours (e.g., “sky” with blue). He reported a semantic gradient in which interference increased as a function of the association between the word and the colour. For instance, the colour-associated word “sky” when printed in an incongruent colour (i.e., red) produces more interference than the colour-neutral word “table”. Thus, strengthening the semantic relationship between the word and colour increased the magnitude of Stroop interference.

Some authors have explained this effect, which arises from compatible and incompatible combinations of colour associated distracters and ink colours, in terms of different mechanisms (Glaser & Glaser, 1989; Klein, 1964; MacKinnon et al., 1985; Posner & Snyder, 1975; Sharma & McKenna, 1998). First, the interference effect produced by incongruent colour associates could reflect early, semantic processes that arises from association between two stimulus dimensions (i.e., word and colour). Concurrent activation of the word and the colour ought to produce a stimulus conflict (Glaser & Glaser, 1989; Risko et al., 2006; Schmidt & Cheesman, 2005). Second, colour associate interference was claimed to be a result of later, response competition processes (Klein, 1964; Sharma & McKenna, 1998). According to this view, a colour-associate is potent enough to affect the output by activating the colour response linked to it, therefore producing response conflict.

Note, however, that all the described accounts are focused on L1 colour associates and their effects on semantic and response processing. Experiment 5 aimed to expand current findings to novel L2 colour associates which were acquired through association with their L1 translations. Therefore, we focused on whether recently trained L2 colour associates produce an interference effect when presented in an incongruent colour; and, if so, does this interference result from early, semantic or late, response competition processes.

## Method

### *Participants*

A total of 141 University of Burgundy undergraduates (130 women, 11 men) were recruited and tested online. The selection criteria were identical as Experiment 3. More demographic information is given in the Results section. Participants received course credit for their participation. The experiments lasted approximately 30 minutes.

### *Apparatus and materials*

The experiment was programmed in the Psytoolkit (Stoet, 2010, 2017) and adapted for running online. As in previous experiments, at the very beginning of the study, participants completed a short survey concerning their language background. The structure of this survey portion was identical to that in Experiment 3. The purpose of this survey was to assure that all participants had the target language dominance (i.e., they are native French speakers) and that they have no previous experience with the Croatian language.

### *Design*

**Learning phase.** The learning phase was identical as in Experiment 3, except for the novel Croatian-French colour associate pairs ("jagoda"-"fraise", "nebo"-"ciel", "deblo"-"bois", and "trava"-"herbe", i.e. strawberry, sky, wood, and grass, respectively). The length of the learning phase and the instructions were unchanged.

**Training phase.** Learning phase was followed by the Training phase. Trial types were the same as in Experiment 4, except the images of colour associates replaced colour boxes (e.g., an image of a strawberry was used instead of a red colour box). The four types of trials in the Training phase were: 1) Croatian target with French labels, 2) Croatian target with image labels, 3) French target with Croatian labels, and 4) Image target with Croatian labels. These trials are illustrated in Figure 17. The number of trials and key response assignment were identical to those in Experiment 4.


***Figure 17.*** Experiment 5 - Types of trials in the training phase

**Stroop task.** The Stroop task was identical as in Experiment 3, with minor modifications. As distracters, four French colour associates (i.e., "fraise", "ciel", "bois", and "herbe") and four recently trained Croatian colour associates (i.e., "jagoda", "nebo", "deblo", and "trava") were used (English translations are strawberry, sky, wood, and grass, respectively). Additionally, to increase the Stroop interference effect, we added four corresponding French colour words (i.e., "rouge"/red, "bleu"/blue, "marron"/brown, and "vert"/green). In total, there were 192 experimental trials across three experimental Stroop task blocks.

***Procedure***

*Learning* and *Training phases* were the same as in Experiment 2. The *Stroop task* portion was the same as in Experiment 4, except for the absence of "catch" trials in the present experiment.

## Results

***Language demographic***

French was a first dominant language (92.20%) and a first language in order of acquisition (97.87%) for most of participants. Participants indicated English (75.89%), Spanish (16.31%), and German (2.13%) as most frequent second dominant languages. As second language in order of acquisition, participants indicated English (82.98%), Spanish (4.26%), Arabic (2.13%), French (1.42%), Italian (1.42%), and Dutch (1.42%). A total of 6.38% indicated a language other than the ones listed above or left the question blank. Almost 95% of participants estimated that their daily exposure to French language is very high (i.e., 81-100%). Mean French language scores are presented in Table 9.

***Table 9.*** Experiment 5 - Mean French language development scores with standard errors

| | Mean | SE |
|---|---|---|
| Acquisition | 1.36 years | .097 |
| Fluent | 3.45 years | .140 |
| Reading | 5.58 years | .086 |
| Fluent reading | 6.84 years | .108 |

***Training phase***

The overall accuracy in the Training phase was high (*MEAN* = 95.53%, *SE* = .227).

**Response times.** Only correct trials were included into RT analysis. Trials with a Croatian target with French labels were responded significantly slower than remaining three types of trial: Croatian target with image labels, $t(140) = 27.887$, $p < .001$, $MEAN_{diff} = 161.853$, $SE_{diff} = 5.80$, *Cohen's d* = 2.35, $BF_{10} > 100$, French target with Croatian labels, $t(140)$

= 12.077, $p < .001$, $MEAN_{diff} = 51.894$, $SE_{diff} = 4.30$, *Cohen's d* = 1.02, $BF_{10} > 100$, and Image target with Croatian labels, $t(140) = 15.907$, $p < .001$, $MEAN_{diff} = 100.937$, $SE_{diff} = 6.35$, *Cohen's d* = 1.34, $BF_{10} > 100$. Participants responded significantly faster on trials with a Croatian target and image targets, relative to the trials with French target and Croatian labels, $t(140) = 22.025$, $p < .001$, $MEAN_{diff} = -109.959$, $SE_{diff} = 4.99$, *Cohen's d* = -1.85, $BF_{10} > 100$, and trials with Image targets and Croatian labels, $t(140) = 11.67$, $p < .001$, $MEAN_{diff} = -60.916$, $SE_{diff} = 5.22$, *Cohen's d* = -.983, $BF_{10} > 100$. Responses on Image target and Croatian labels were significantly faster than responses on French target and Croatian labels, $t(140) = 9.974$, $p < .001$, $MEAN_{diff} = 49.043$, $SE_{diff} = 4.92$, *Cohen's d* = .84, $BF_{10} > 100$.

**Percentage error.** Participants had significantly higher percentage error on trials with a Croatian target and French labels, relative to trials with a Croatian target and image labels, $t(140) = 7.529$, $p < .001$, $MEAN_{diff} = 1.990$, $SE_{diff} = .264$, *Cohen's d* = .634, $BF_{10} > 100$, French target and Croatian labels, $t(140) = 10.59$, $p < .001$, $MEAN_{diff} = 2.391$, $SE_{diff} = .226$, *Cohen's d* = .892, $BF_{10} > 100$, and Image target with Croatian labels, $t(140) = 14.169$, $p < .001$, $MEAN_{diff} = 3.401$, $SE_{diff} = .240$, *Cohen's d* = 1.19, $BF_{10} > 100$. Responses on trials with an Image target and Croatian labels had lower percentage errors relative to trials with a Croatian target and image labels, $t(140) = 7.525$, $p < .001$, $MEAN_{diff} = 1.410$, $SE_{diff} = .187$, *Cohen's d* = .634, $BF_{10} > 100$, and French target with Croatian labels, $t(140) = 5.514$, $p < .001$, $MEAN_{diff} = 1.009$, $SE_{diff} = .183$, *Cohen's d* = .464, $BF_{10} > 100$. Response times and percentage errors with standard errors for all training trial types are presented in Table 10.

***Table 10.*** Experiment 5 - Response times and percentage errors with standard errors in the training phase

| Type of trial | Response Time | | Percentage Error | |
|---|---|---|---|---|
| | Mean | SE | Mean | SE |
| Croatian target, French labels | 1132.75 | 10.7 | 6.31 | .30 |
| Croatian target, image labels | 970.9 | 9.89 | 4.32 | .23 |
| French target, Croatian labels | 1080.85 | 10.69 | 3.92 | .24 |
| Image target, Croatian labels | 1031.81 | 11.95 | 2.91 | .20 |

***Stroop task***

**Response time.** There was a main effect of Congruency, $F(2,280) = 26.101$, $p < .001$, $MSE = 2489.257$, $\eta^2_p = .157$, $BF_{10} > 100$, indicating a difference in response speed between different types of trial (i.e. identity/same response/different response). The main effect of Word Type was observed, $F(2,280) = 6.576$, $p < .01$, $MSE = 1944.128$, $\eta^2_p = .045$, $BF_{10} = 1.55$, indicating a difference in response times between word types (i.e. French colour associates/Croatian colour associates/French colour words). The interaction between Congruency and Word Type was also significant, $F(4,560) = 19.046$, $p < .001$, $MSE = 2119.955$, $\eta^2_p = .120$, $BF_{10} > 100$.

Comparisons were conducted on each word type separately with results displayed in Figure 18. For French colour associates, neither stimulus conflict effect, $t(140) = 1.394$, $p = .165$, $MEAN_{diff} = -7.625$, $SE_{diff} = 5.47$, *Cohen's d* = -.117, $BF_{10} = .242$, nor response conflict effect, $t(140) = .459$, $p > .05$, $MEAN_{diff} = -2.43$, $SE_{diff} = 5.29$, *Cohen's d* = -.039, $BF_{10} = .104$, were significant. The overall Stroop interference effect, $t(140) = 1.762$, $p = .08$, $MEAN_{diff} = -10.055$, $SE_{diff} = 5.71$, *Cohen's d* = -.148, $BF_{10} = .422$, was only marginally significant. Similarly for Croatian colour associates, the stimulus conflict effect, $t(140) = .034$, $p > .05$, $MEAN_{diff} = .195$, $SE_{diff} = 5.73$, *Cohen's d* = .003, $BF_{10} = .094$, response conflict effect, $t(140) = .169$, $p > .05$, $MEAN_{diff} = -.802$, $SE_{diff} = 4.75$, *Cohen's d* = -.014, $BF_{10} = .095$, and overall Stroop interference effect, $t(140) = .110$, $p > .05$, $MEAN_{diff} = -0.607$, $SE_{diff} = 5.53$, *Cohen's d* =

-.009, $BF_{10}$ = .094, did not reach statistical significance. As expected, for French colour words, both the stimulus conflict effect, $t(140) = 7.203$, $p < .001$, $MEAN_{diff}$ = -44.329, $SE_{diff}$ = 6.15, *Cohen's d* = -.607, $BF_{10} > 100$, and response conflict effect, $t(140) = 2.615$, $p = .01$, $MEAN_{diff}$ = -17.124, $SE_{diff}$ = 6.55, *Cohen's d* = -.220, $BF_{10}$ = 2.48, were significant, as well as overall Stroop interference, $t(140) = 11.383$, $p < .001$, $MEAN_{diff}$ = -61.453, $SE_{diff}$ = 5.40, *Cohen's d* = -.959, $BF_{10} > 100$.

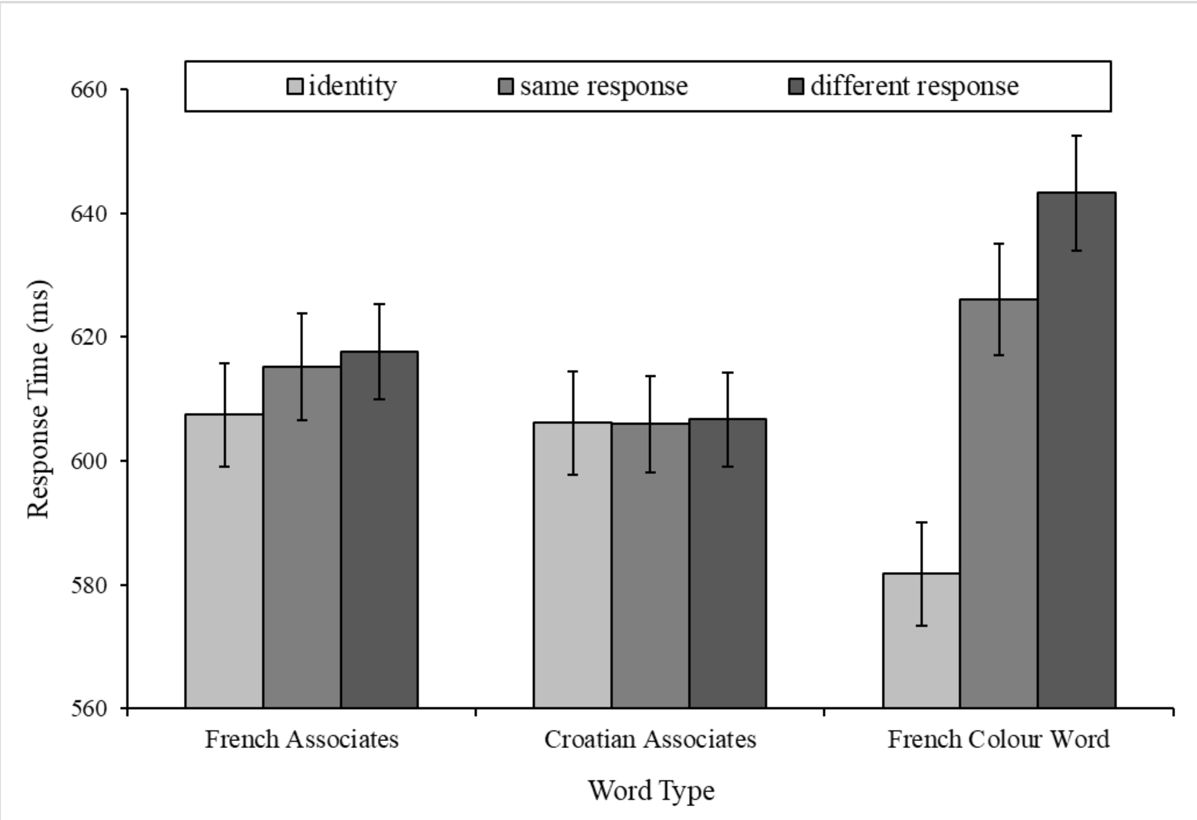


***Figure 18.*** Experiment 5 - Response times with standard errors for French associates, Croatian associates, and French colour words in the Stroop task

There was no evidence for any differences in the magnitude of the stimulus conflict effect between French colour associates and Croatian colour associates, $t(140) = 1.116$, $p > .05$, $MEAN_{diff}$ = 7.820, $SE_{diff}$ = 7.01, *Cohen's d* = .094, $BF_{10}$ = .172. The magnitude of the stimulus conflict effect was significantly larger for French colour words relative to French colour associates, $t(140) = 4.456$, $p < .001$, $MEAN_{diff}$ = -36.704, $SE_{diff}$ = 8.24, *Cohen's d* = -

.375, $BF_{10} > 100$, and Croatian colour associates, $t(140) = 5.167$, $p < .001$, $MEAN_{diff} = -44.524$, $SE_{diff} = 8.62$, *Cohen's d* = -.435, $BF_{10} > 100$.

There was no evidence for any differences in the magnitude of the response conflict effect between French colour associates and Croatian colour associates, $t(140) = .244$, $p > .05$, $MEAN_{diff} = 1.628$, $SE_{diff} = 6.67$, *Cohen's d* = .021, $BF_{10} = .096$, nor between French colour associates and French colour words, $t(140) = -1.756$, $p > .05$, $MEAN_{diff} = -14.694$, $SE_{diff} = 8.37$, *Cohen's d* = -.148, $BF_{10} = .418$. The magnitude of the response conflict effect was significantly larger for French colour words relative to Croatian colour associates, $t(140) = 2.205$, $p < .05$, $MEAN_{diff} = -16.322$, $SE_{diff} = 7.40$, *Cohen's d* = -.186, $BF_{10} = .978$. There was no difference in the magnitude of Stroop interference between French colour associates and Croatian colour associates, $t(140) = 1.261$, $p > .05$, $MEAN_{diff} = 9.448$, $SE_{diff} = 7.49$, *Cohen's d* = .106, $BF_{10} = .203$. However, Stroop interference was significantly larger for French colour words relative to French colour associates, $t(140) = 6.434$, $p < .001$, $MEAN_{diff} = -51.398$, $SE_{diff} = 7.99$, *Cohen's d* = -.542, $BF_{10} > 100$, and Croatian colour associates, $t(140) = 7.806$, $p < .001$, $MEAN_{diff} = -60.846$, $SE_{diff} = 7.79$, *Cohen's d* = -.657, $BF_{10} > 100$.

**Percentage error.** The main effects of Congruency, $F(2,280) = 11.228$, $p < .001$, $MSE = 22.093$, $\eta^2_p = .074$, $BF_{10} > 100$, and Word Type, $F(2,280) = 3.602$, $p < .05$, $MSE = 16.849$, $\eta^2_p = .025$, $BF_{10} = .179$, were observed. The interaction between Congruency and Word Type was significant, $F(4,560) = 6.016$, $p < .001$, $MSE = 18.911$, $\eta^2_p = .041$, $BF_{10} = 82.766$.

Comparisons were conducted on each word type separately with results displayed at Figure 19. For French colour associates, the stimulus conflict effect was marginally significant in the reversed direction with a higher percentage error on identity than on same response trials, $t(140) = 1.943$, $p = .054$, $MEAN_{diff} = 1.161$, $SE_{diff} = .598$, *Cohen's d* = .164, $BF_{10} = .582$. The response conflict effect, $t(140) = .566$, $p > .05$, $MEAN_{diff} = -.303$, $SE_{diff} =$

.536, *Cohen's d* = .048, $BF_{10}$ = .11, and overall Stroop interference effect, $t(140) = 1.668$, $p > .05$, $MEAN_{diff}$ = .858, $SE_{diff}$ = .514, *Cohen's d* = .141, $BF_{10}$ = .362, were not significant. For Croatian colour associates, only the response conflict effect was significant, $t(140) = 2.521$, $p = .013$, $MEAN_{diff}$ = -1.182, $SE_{diff}$ = .469, *Cohen's d* = -.212, $BF_{10}$ = 1.97, while the stimulus conflict effect, $t(140) = 1.106$, $p > .05$, $MEAN_{diff}$ = .628, $SE_{diff}$ = .568, *Cohen's d* = .093, $BF_{10}$ = .17, and overall Stroop interference, $t(140) = 1.232$, $p > .05$, $MEAN_{diff}$ = -.554, $SE_{diff}$ = .45, *Cohen's d* = -.104, $BF_{10}$ = .196, did not reach statistical significance. For French colour words, there was no stimulus conflict effect, $t(140) = 1.326$, $p > .05$, $MEAN_{diff}$ = .684, $SE_{diff}$ = .516, *Cohen's d* = .112, $BF_{10}$ = .221, but there was a significant response conflict effect, $t(140) = 5.692$, $p < .001$, $MEAN_{diff}$ = -3.104, $SE_{diff}$ = .545, *Cohen's d* = -.479, $BF_{10}$ > 100, and significant Stroop interference effect, $t(140) = 4.199$, $p < .001$, $MEAN_{diff}$ = -2.420, $SE_{diff}$ = .576, *Cohen's d* = -.354, $BF_{10}$ = 324.

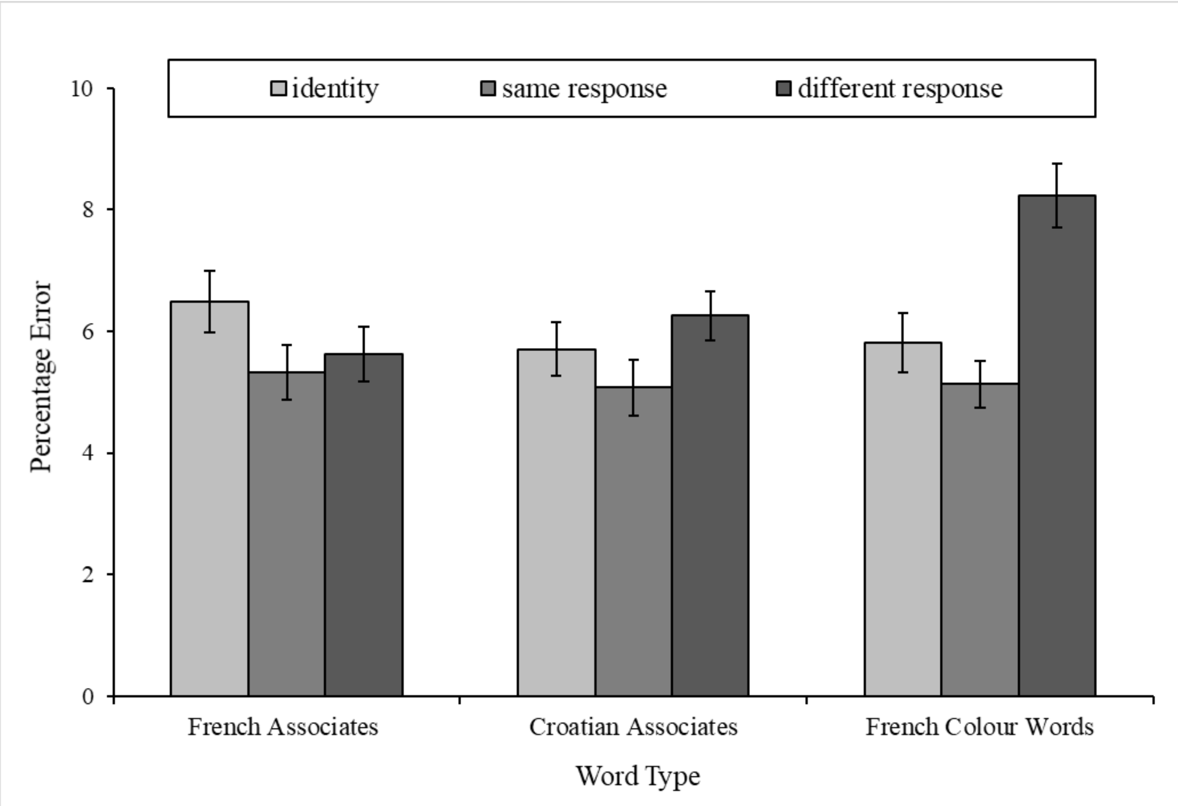


***Figure 19.*** Experiment 5 - Percentage errors with standard errors for French associates, Croatian associates, and French colour words in the Stroop task

There was no evidence for any difference in the magnitude of the stimulus conflict effect between French colour associates and Croatian colour associates, $t(140) = .655$, $p > .05$, $MEAN_{diff} = -.533$, $SE_{diff} = .813$, *Cohen's d* = -.055, $BF_{10} = .116$, French colour associates and French colour words, $t(140) = .647$, $p > .05$, $MEAN_{diff} = -.477$, $SE_{diff} = .737$, *Cohen's d* = -.054, $BF_{10} = .115$, nor Croatian colour associates and French colour words, $t(140) = .075$, $p > .05$, $MEAN_{diff} = .056$, $SE_{diff} = .744$, *Cohen's d* = .006, $BF_{10} = .094$.

No difference in the magnitude of the response conflict effect between French colour associates and Croatian colour associates, $t(140) = 1.196$, $p > .05$, $MEAN_{diff} = -.879$, $SE_{diff} = .735$, *Cohen's d* = -.101, $BF_{10} = .188$, was observed. The magnitude of the response conflict effect was significantly larger for French colour words relative to French colour associates, $t(140) = 3.852$, $p < .001$, $MEAN_{diff} = -2.801$, $SE_{diff} = .727$, *Cohen's d* = -.324, $BF_{10} = 94.9$, and Croatian colour associates, $t(140) = 2.825$, $p < .01$, $MEAN_{diff} = -1.922$, $SE_{diff} = .680$, *Cohen's d* = -.238, $BF_{10} = 4.22$.

Stroop interference was larger for Croatian colour associates than for French colour associates, $t(140) = 2.273$, $p < .05$, $MEAN_{diff} = -1.412$, $SE_{diff} = .621$, *Cohen's d* = -.191, $BF_{10} = 1.13$. The magnitude of Stroop interference was larger for French colour words than for French colour associates, $t(140) = 4.394$, $p < .001$, $MEAN_{diff} = -3.278$, $SE_{diff} = .746$, *Cohen's d* = -.37, $BF_{10} > 100$, and Croatian colour associates, $t(140) = 2.418$, $p < .05$, $MEAN_{diff} = -1.866$, $SE_{diff} = .772$, *Cohen's d* = -.204, $BF_{10} = 1.56$.

**Discussion**

When colour associates are mentioned in a linguistic context, they activate their corresponding colour representations (Tanaka et al., 2001). However, it has not been investigated whether colour associates, as words that have been historically related to colours, could facilitate L2 acquisition. In Experiment 5, we trained participants with novel Croatian colour associates associated with their French translations (e.g., “nebo”- “ciel”, Croatian and French translation of “sky”, respectively”). Native language associates printed in incongruent colours interfere with colour identification (Risko et al., 2006; Schmidt & Cheesman, 2005). However, in the present experiment, only marginally significant Stroop interference was observed for French colour associates. Even though the response latencies showed a tendency toward a stimulus conflict effect, it failed to reach significance.

Apart from French colour associates themselves, we tested the source of the interference effect for Croatian colour associates learned through the association with their French equivalent. Therefore, novel Croatian words have never been directly related to their underlying colour concept. No significant interference effect (neither for the response times nor percentage errors) was observed for Croatian colour associates. Thus, the idea of decomposing this interference into stimulus and response conflict subcomponents was not supported. However, data inspection revealed the significant response conflict for Croatian (L2) colour associates in errors. Interestingly, this response conflict effect was larger (although not significantly) than French associates response conflict.

As expected, the most pronounced effects were evidenced for French colour words. This is not surprising given that L1 strongly influences semantic and response processing. The purpose of involving French colour words was to increase the interference effect for novel words associated with these colours, similarly as in work of Geukes and colleagues (2015). In their experiment, novel, to-be-learned words were associated with native colour words. The interference effect was observed for novel words in the Stroop task administered immediately after training, but only with the presence of native colour words (Geukes et al., 2015). However, the presence of L1 colour words in the Stroop task could not boost the interference effect in novel L2 words since they were not directly associated.

To sum up, our intention to investigate a source of the interference effect in novel L2 colour associated words failed due to the lack of an L2 interference effect. A few possible explanations of the observed pattern of results should be considered. For instance, it is possible that our training procedure was not appropriate for acquisition of this type of L2 words (i.e., L2 colour associates could not be acquired/trained this way). Other possible explanations for the obtained pattern of results are that the administered training was too short, or that participants simply did not pay attention to word distracters in the Stroop task.

## Experiment 6

Experiment 6 investigated further an L2 associate interference effect, with a slightly modified Stroop task procedure relative to the one used in Experiment 5. As already discussed (see Experiment 5), learning novel L2 colour associates in combination with their L1 translation did not result in an L2 interference effect in a subsequent Stroop task. One issue with the lack of substantial effects for L1 and L2 colour associates in Experiment 5 was that participants might not have assigned sufficient attention to the distracters when responding to the ink colour (Besner et al., 1997; Manwell et al., 2004). Therefore, to force them to assign more attention, we decided to replicate Experiment 5 with the addition of “catch” trials, similarly to Experiments 3 and 4.

### Method

#### *Participants*

A total of 92 University of Burgundy undergraduates (86 women, 6 men) were recruited online. The experiment lasted approximately 30 minutes, and in return participants were given credits. The selection criteria were identical as Experiment 5. Detailed demographic information about the sample can be found in the Results section. None of participants of participants had prior Croatian knowledge or had studied Croatian in the school. Also, none of them were able to translate the given Croatian colour words.

### *Apparatus and materials, design, and procedure*

The experiment was identical in all respects to Experiment 5, with the following exceptions. Catch trials were added to the Stroop task, which was otherwise identical. In particular, two novel Croatian words ("mokar" and "petak") served as catch trials on which participants had to withhold their response. The timing for catch trials was the same as in Experiment 4.

## Results

### *Language demographic*

As expected, vast majority of participants indicated French as their first languages in order of dominance (90.22%) and order of acquisition (95.65%). Participants rated English (70.65%), Spanish (14.13%), German (6.52%), and Turkish (2.17%) as their second languages in order of dominance. English was second language in order of acquisition for most participants (77.17%), followed by Spanish (7.61%) and German (5.43%). Participants (85.87%) estimated that they are exposed to French language for the majority of their time (81-100%) on a daily basis. Average French language development scores are presented in Table 11.

***Table 11.*** Experiment 6 - Mean French language development scores with standard errors

| | Mean | SE |
|---|---|---|
| Acquisition | 1.56 years | .164 |
| Fluent | 3.67 years | .210 |
| Reading | 5.46 years | .105 |
| Fluent reading | 6.75 years | .145 |

### *Training phase*

The overall accuracy in the Training phase was high (*MEAN* = 94.78%, *SE* = .322).

**Response time.** Only correct trials were included in response speed analysis. Responses in trials with a Croatian target and French labels were slower than in three other types of trial: Croatian target and image labels, $t(91) = 23.232$, $p < .001$, $MEAN_{diff} = 168.829$, $SE_{diff} = 7.27$, *Cohen's d* = 2.42, $BF_{10} > 100$, French target and Croatian labels, $t(91) = 10.576$, $p < .001$, $MEAN_{diff} = 57.993$, $SE_{diff} = 5.48$, *Cohen's d* = 1.1, $BF_{10} > 100$, and Image target and Croatian labels, $t(91) = 13.084$, $p < .001$, $MEAN_{diff} = 101.01$, $SE_{diff} = 7.72$, *Cohen's d* = 1.36, $BF_{10} > 100$. Responses on trials with a Croatian target and image labels were faster than responses on trials with French target and Croatian labels, $t(91) = 15.064$, $p < .001$, $MEAN_{diff} = -110.835$, $SE_{diff} = 7.36$, *Cohen's d* = -1.57, $BF_{10} > 100$, and on trials with an Image target and Croatian labels, $t(91) = 10.838$, $p < .001$, $MEAN_{diff} = -67.819$, $SE_{diff} = 6.26$, *Cohen's d* = -1.13, $BF_{10} > 100$. Trials with an Image target and Croatian labels were responded faster than trials with a French target and Croatian labels, $t(91) = 6.243$, $p < .001$, $MEAN_{diff} = 43.017$, $SE_{diff} = 6.89$, *Cohen's d* = .651, $BF_{10} > 100$.

**Percentage error.** Percentage error was significantly higher for trials with a Croatian target and French labels, relative to another types of trial: Croatian target and image labels, $t(91) = 6.972$, $p < .001$, $MEAN_{diff} = 2.101$, $SE_{diff} = .301$, *Cohen's d* = .727, $BF_{10} > 100$, French target and Croatian labels, $t(91) = 9.729$, $p < .001$, $MEAN_{diff} = 2.955$, $SE_{diff} = .304$, *Cohen's d* = 1.01, $BF_{10} > 100$, and Image target with Croatian labels, $t(91) = 9.561$, $p < .001$, $MEAN_{diff} = 3.077$, $SE_{diff} = .322$, *Cohen's d* = .997, $BF_{10} > 100$. Responses on trials with a Croatian target and image labels were less accurate than responses on trials with a French target and Croatian labels, $t(91) = 2.832$, $p = .01$, $MEAN_{diff} = .854$, $SE_{diff} = .302$, *Cohen's d* = .295, $BF_{10} = 4.84$, and trials with an Image target and Croatian labels, $t(91) = 3.3$, $p = .001$, $MEAN_{diff} = .976$, $SE_{diff} = .296$, *Cohen's d* = .344, $BF_{10} = 17.1$. The response times and percentage error with standard error for all types of trial are presented in Table 12.

***Table 12.*** Experiment 6 - Response times and percentage errors with standard errors in the training phase

| Type of trial | Response Time | | Percentage Error | |
|---|---|---|---|---|
| | Mean | SE | Mean | SE |
| Croatian target, French labels | 1129.77 | 14.84 | 7.25 | .44 |
| Croatian target, image labels | 960.94 | 12.50 | 5.15 | .40 |
| French target, Croatian labels | 1071.78 | 14.10 | 4.29 | .30 |
| Image target, Croatian labels | 1028.76 | 13.99 | 4.17 | .32 |

***Stroop task***

The performance on catch trials was analysed separately from experimental trials within the Stroop tasks. On average, catch trials accounted for approximately 14.23% of experimental trials (*MEAN* = 44.41, *SE* = .147). Participants were relatively successful in withholding their responses on catch trials (*MEAN* = 82.92%, *SE* = 1.332).

**Response time.** The correct response time data are presented in Figure 20. To analyse response times, a Congruency (identity/same response/different response) by Word Type (French colour associates/Croatian colour associates/French colour word) within-subject repeated-measures analysis of variance was conducted. The main effect of Congruency, $F(2,182) = 43.594$, $p < .001$, $MSE = 3127.642$, $\eta^2p = .324$, $BF_{10} > 100$ and Word Type, $F(2,182) = 8.762$, $p < .001$, $MSE = 2783.539$, $\eta^2p = .088$, $BF_{10} = 8.95$, were observed. The interaction between Congruency and Word Type was significant, $F(4,364) = 11.832$, $p < .001$, $MSE = 3183.782$, $\eta^2p = .115$, $BF_{10} > 100$.

Comparison on each Word Type were separately conducted. For French colour associates, there was a significant stimulus conflict effect (same response – identity), $t(91) = 2.469$, $p = .01$, $MEAN_{diff} = -19.513$, $SE_{diff} = 7.9$, *Cohen's d* = -.257, $BF_{10} = 2.03$, response conflict effect (different response – same response), $t(91) = 2.462$, $p = .01$, $MEAN_{diff} = -15.582$, $SE_{diff} = 6.33$, *Cohen's d* = -.257, $BF_{10} = 2.00$, and overall Stroop interference effect (different response – identity), $t(91) = 4.539$, $p < .001$, $MEAN_{diff} = -35.094$, $SE_{diff} = 7.73$, *Cohen's d* = -.473, $BF_{10}$ > 100. For Croatian colour associates, there was no significant stimulus conflict effect (same response – identity), $t(91) = .232$, $p > .05$, $MEAN_{diff} = 2.109$, $SE_{diff} = 9.11$, *Cohen's d* = .024, $BF_{10} = .118$, no response conflict effect (different response – same response), $t(91) = 1.505$, p > .05, $MEAN_{diff} = -12.229$, $SE_{diff} = 8.12$, *Cohen's d* = -.157, $BF_{10} = .342$. The overall Stroop interference effect (different response – identity) also failed to reach significance, $t(91) = 1.416$, $p > .05$, $MEAN_{diff} = -10.12$, $SE_{diff} = 7.15$, *Cohen's d* = -.148, $BF_{10} = .302$. For French colour words, we observed a significant stimulus conflict effect (same response – identity), $t(91) = 4.653$, $p < .001$, $MEAN_{diff} = -48.591$, $SE_{diff} = 10.4$, *Cohen's d* = -.485, $BF_{10}$ > 100, response conflict effect (different response – same response), $t(91) = 4.985$, $p < .001$, $MEAN_{diff} = -39.549$, $SE_{diff} = 7.93$, *Cohen's d* = -.52, $BF_{10}$ > 100, and overall Stroop interference effect (different response – identity), $t(91) = 9.56$, $p < .001$, $MEAN_{diff} = -88.141$, $SE_{diff} = 9.22$, *Cohen's d* = -.997, $BF_{10}$ > 100.

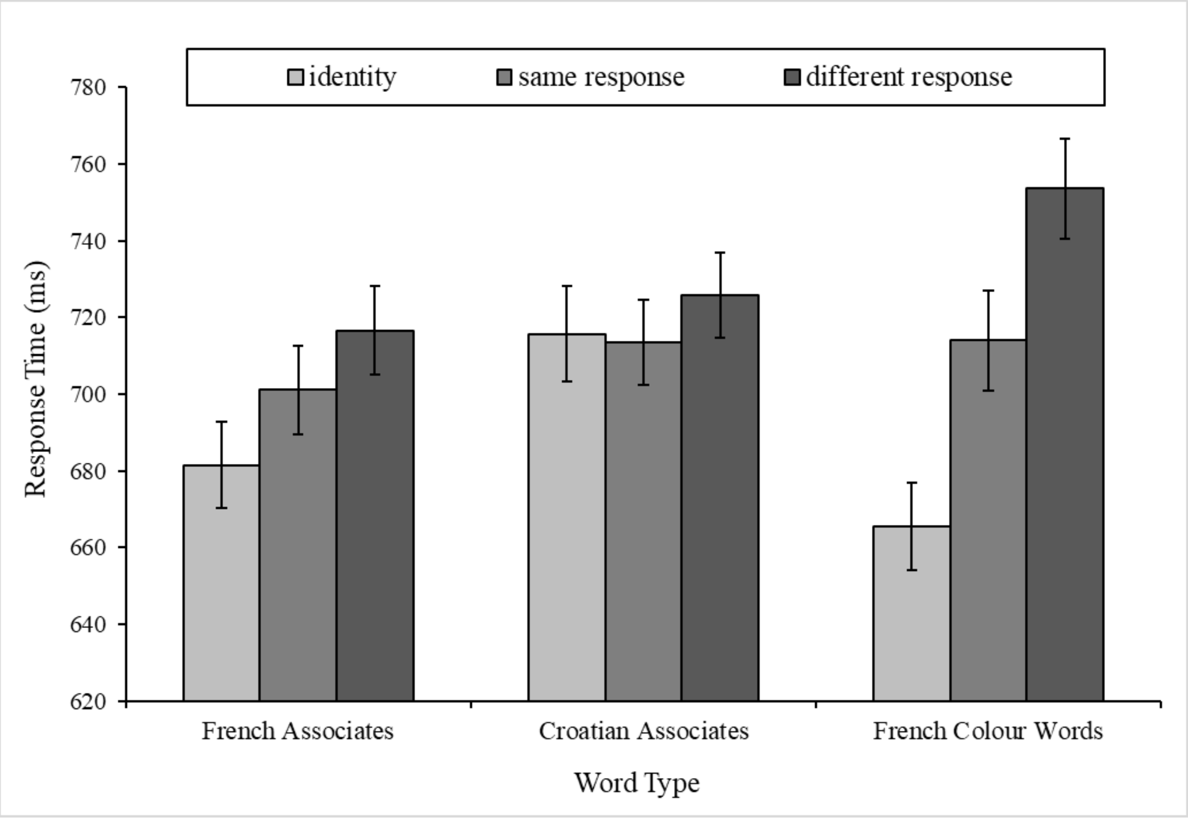


***Figure 20.*** Experiment 6 - Response times with standard errors for French associates, Croatian associates, and French colour words in the Stroop task

The difference in the magnitude of the stimulus conflict effect between French colour associates and Croatian colour associates was marginally significant, $t(91) = 1.82$, $p = .07$, $MEAN_{diff} = 21.621$, $SE_{diff} = 11.9$, *Cohen's d* = .19, $BF_{10} = .559$. The magnitude of the stimulus conflict effect was significantly larger for French colour words relative to French colour associates, $t(91) = 2.166$, $p < .05$, $MEAN_{diff} = -29.079$, $SE_{diff} = 13.4$, *Cohen's d* = -.226, $BF_{10}$ = 1.06, and Croatian colour associates, $t(91) = 3.574$, $p < .001$, $MEAN_{diff} = -50.7$, $SE_{diff} = 14.2$, *Cohen's d* = -.373, $BF_{10} = 38.6$. There was no evidence for any difference in the magnitude of the response conflict effect between French colour associates and Croatian colour associates, $t(91) = .336$, $p > .05$, $MEAN_{diff} = 3.352$, $SE_{diff} = 9.99$, *Cohen's d* = .035, $BF_{10} = .122$. The response conflict effect was significantly larger for French colour words relative to French colour associates, $t(91) = 2.410$, $p < .05$, $MEAN_{diff} = -23.968$, $SE_{diff} = 9.94$, *Cohen's d* = -.251, $BF_{10} = 1.78$, and Croatian colour associates, $t(91) = 2.887$, $p < .01$, $MEAN_{diff} = -27.32$, $SE_{diff}$ = 9.46, *Cohen's d* = -.301, $BF_{10} = 5.57$. Stroop interference was significantly larger for French colour associates than for Croatian colour associates, $t(91) = 2.298$, $p < .05$, $MEAN_{diff}$ = 24.975, $SE_{diff} = 10.9$, *Cohen's d* = .24, $BF_{10} = 1.40$. The magnitude of Stroop interference was larger for French colour words than for French colour associates, $t(91) = 4.226$, $p < .001$, $MEAN_{diff} = -53.046$, $SE_{diff} = 12.6$, *Cohen's d* = -.441, $BF_{10} > 100$, and Croatian colour associates, $t(91) = 6.181$, $p < .001$, $MEAN_{diff} = -78.021$, $SE_{diff} = 12.6$, *Cohen's d* = -.644, $BF_{10}$ > 100.

**Percentage error.** The percentage error data are presented in Figure 21. Again, a Congruency (identity/same response/different response) by Word Type (French colour associates/Croatian colour associates/French colour word) within-subject repeated-measures analysis of variance was conducted. The main effect of Congruency was significant, $F(2,182)$ = 16.879, $p < .001$, $MSE = 33.939$, $\eta^2 p = .156$, $BF_{10} > 100$. However, there was no main effect of Word Type, $F(2,182) = .274$, $p > .05$, $MSE = 28.3$, $\eta^2 p = .003$, $BF_{10} = .017$. Importantly, the interaction between Congruency and Word Type was significant, $F(4,364) = 7.386$, $p <$ .001, $MSE = 30.692$, $\eta^2p = .075$, $BF_{10} > 100$.

We conducted planned comparisons on each word type separately. For French colour associates, there was neither a stimulus conflict effect (same response – identity), $t(91) = .618$, $p > .05$, $MEAN_{diff} = .492$, $SE_{diff} = .797$, *Cohen's d* = .064, $BF_{10} = .139$, nor a response conflict effect (different response – same response), $t(91) = 1.451$, $p > .05$, $MEAN_{diff} = -1.118$, $SE_{diff} =$ .77, *Cohen's d* = -.151, $BF_{10} = .317$. The overall Stroop interference effect (different response – identity) also failed to reach significance, $t(91) = .804$, $p > .05$, $MEAN_{diff} = -.626$, $SE_{diff} =$ .778, *Cohen's d* = -.084, $BF_{10} = .158$. For Croatian colour associates, there was no stimulus conflict effect, $t(91) = 1.598$, $p > .05$, $MEAN_{diff} = 1.542$, $SE_{diff} = .965$, *Cohen's d* = .167, $BF_{10}$ = .392, but there was a significant response conflict effect, $t(91) = 2.557$, $p = .01$, $MEAN_{diff} = -$ 1.936, $SE_{diff} = .757$, *Cohen's d* = -.267, $BF_{10} = 2.48$. The overall Stroop interference effect was not significant, $t(91) = .43$, $p > .05$, $MEAN_{diff} = -.394$, $SE_{diff} = .916$, *Cohen's d* = -.045, $BF_{10} = .126$. For French colour words, there was no stimulus conflict effect, $t(91) = 1.04$, $p >$ .05, $MEAN_{diff} = -.815$, $SE_{diff} = .784$, *Cohen's d* = -.108, $BF_{10} = .194$, but there was a response conflict effect, $t(91) = 5.899$, $p < .001$, $MEAN_{diff} = -4.967$, $SE_{diff} = .842$, *Cohen's d* = -.615, $BF_{10} > 100$, and overall interference effect, $t(91) = 6.834$, $p < .001$, $MEAN_{diff} = -5.782$, $SE_{diff} =$ .846, *Cohen's d* = -.712, $BF_{10} > 100$.

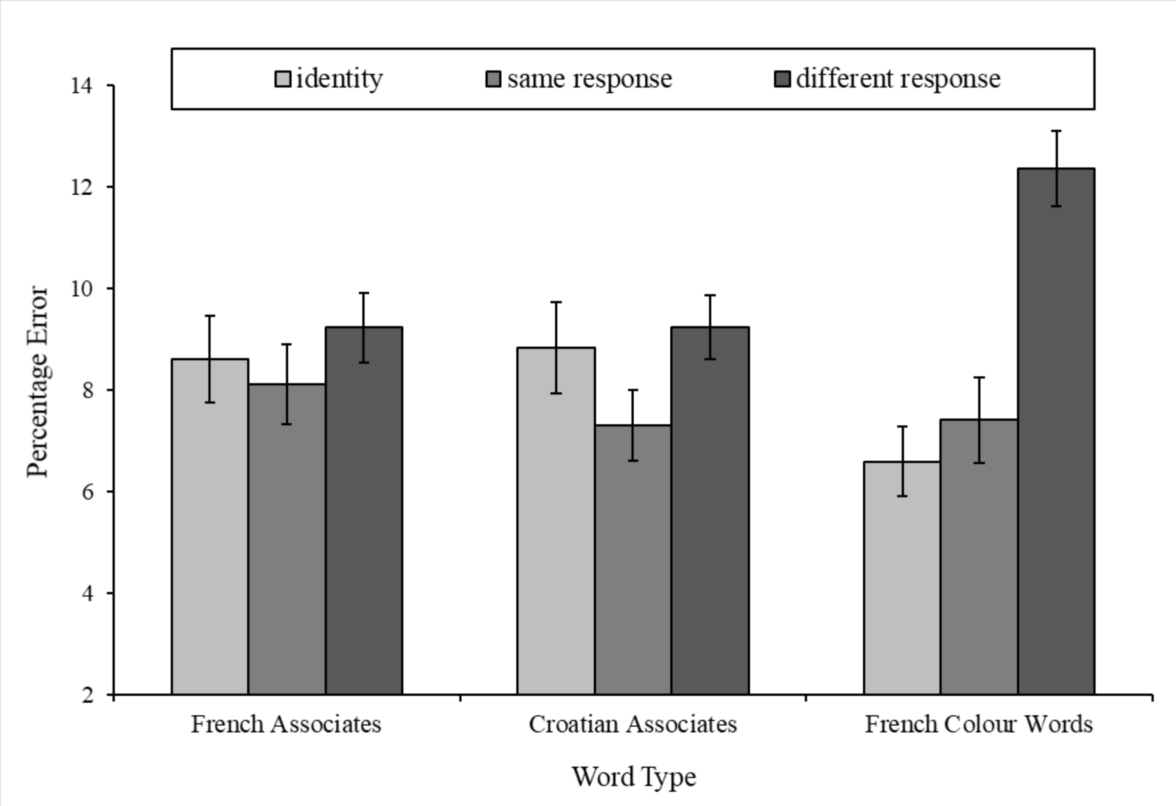


***Figure 21.*** Experiment 6 - Percentage errors for French associates, Croatian associates, and French colour words in the Stroop task

There was no evidence for any difference in the magnitude of the stimulus conflict effect between French colour associates and Croatian colour associates, $t(91) = .89$, $p > .05$, $MEAN_{diff} = 1.05$, $SE_{diff} = 1.18$, *Cohen's d* = .093, $BF_{10} = .169$, French colour associates and French colour words, $t(91) = 1.082$, $p > .05$, $MEAN_{diff} = -1.308$, $SE_{diff} = 1.21$, *Cohen's d* = -.113, $BF_{10} = .203$, nor Croatian colour associates and French colour words, $t(91) = 1.732$, $p = .09$, $MEAN_{diff} = -2.358$, $SE_{diff} = 1.36$, *Cohen's d* = -.181, $BF_{10} = .484$. No difference in the magnitude of response conflict effect between French colour associates and Croatian colour associates was observed, $t(91) = .783$, $p > .05$, $MEAN_{diff} = -.817$, $SE_{diff} = 1.04$, *Cohen's d* = -.082, $BF_{10} = .155$. The magnitude of response conflict effect was significantly larger for French colour words relative to French colour associates, $t(91) = 4.048$, $p < .001$, $MEAN_{diff} = -3.848$, $SE_{diff} = .951$, *Cohen's d* = -.422, $BF_{10} > 100$, and Croatian colour associates, $t(91) = 2.77$, $p = .01$, $MEAN_{diff} = -3.031$, $SE_{diff} = 1.09$, *Cohen's d* = -.289, $BF_{10} = 4.13$. There was no difference in the magnitude of Stroop interference between French colour associates and

Croatian colour associates, $t(91) = .204$, $p > .05$, $MEAN_{diff} = .232$, $SE_{diff} = 1.14$, *Cohen's d* = .021, $BF_{10} = .118$. However, Stroop interference was significantly larger for French colour words than for French colour associates, $t(91) = 4.886$, $p < .001$, $MEAN_{diff} = -5.156$, $SE_{diff} = 1.06$, *Cohen's d* = -.509, $BF_{10} > 100$, and Croatian colour associates, $t(91) = 4.12$, $p < .001$, $MEAN_{diff} = -5.388$, $SE_{diff} = 1.31$, *Cohen's d* = -.43, $BF_{10} > 100$.

**Discussion**

Experiment 6 conceptually replicated Experiment 5 with the addition of "catch" trials, that forced participants to attend the written distracters. The effects of native language colour associates have previously been restricted to a stimulus conflict effect (Schmidt & Cheesman, 2005), suggesting that the relationship of a colour associate and ink colour is semantic in nature (Glaser & Glaser, 1989; MacKinnon et al., 1985; Stirling, 1979). This stimulus conflict effect had been confirmed by Experiment 6. For instance, the word "sky" activates the corresponding colour concept "blue", which interferes with the encoding of target colour on incongruent trials. Interestingly, we observed also a significant response conflict effect for French colour associates, which was not the case in some previous studies (Experiment 5, see also Schmidt & Cheesman, 2005). The difference between same response and different response trials was significant for response latency but not for percentage error data. This finding could seem surprising considering the lack of a direct link between the response elicited by an ink colour (one of the keys) and the response elicited by a colour associate (in our experiment, none). Therefore, it is possible that native language colour associates could prime motor responses. To do so, the processed colour associated needs to prime the corresponding colour concept before it generates a motor response. Another explanation for these pronounced colour associate interference effects could be the presence of the native language colour words in the Stroop task (Geukes et al., 2015). For instance, colour associates have been strongly related to colours in the response set (Klein, 1964). This role of response set membership has been clearly established for colour words (Klein, 1964; Proctor, 1978; Sharma & McKenna, 1998), but also for colour associates (Risko et al., 2006). Colour associates interfered with colour identification more than neutral words when they are related to colours in the response set. The presence of corresponding native colour words in the Stroop task that match key response labels could increase interference effects in colour associates. For instance, the word “ciel” (sky in French) should produce interference if the display colour blue is in the response set. The presence of “bleu” (blue in French) both as a distracter and potential response possibly spreads activation (Collins & Loftus, 1975) to other, closely related semantic concepts (i.e., “ciel”).

Apart from the L1 associates’ effect, the main hypotheses were oriented toward their role in on L2 acquisition. Of our particular interest were interference effects for recently trained Croatian colour associates. The overall Stroop interference effect as well as its components failed to reach significance, with a pattern of results that hinted at a response conflict effect. Both response time and percentage error data for L2 colour associates was numerically in the expected direction. Moreover, similarly as in Experiment 5, Croatian colour associates produced a significant response conflict in the errors. The nature of training could possibly explain these findings, independently of its length. For instance, training trials required participants to match novel Croatian colour associates with its French translation (or reverse) or corresponding picture (or reverse). When considering the distinction between lexical and semantic levels of word representations (Kroll & Stewart, 1994), this training procedure required both lexical and semantic processing. Lexical processing results from connection between L1 and L2 word forms (i.e., Croatian colour associate paired with its corresponding French colour associate or vice versa) that has been reinforced during training. Therefore, this hint toward a response conflict effect in Croatian colour associates is based on the translations of these colour associates into their corresponding colour associate. On the other hand, a lack of semantic processing for Croatian colour associates could be explained by weak connection with underlying semantic representation. In the training phase, Croatian colour associates have been associated with corresponding picture, but never directly with a colour word, colour patch, or have been coloured itself. In other words, the presence of stimulus conflict for novel Croatian colour associates could possibly arise with an adapted training procedure with more explicit connections between novel words and conceptual representations.

To sum up, Experiment 6 aimed to extend the scope of the learning context by introducing colour associates (rather than only colour words) in the learning procedure. However, the lack of substantial effects for novel L2 colour associates could either reflect the inappropriateness of this training type in acquiring L2 words or a need for further adaptation of the current word training.

## General discussion

The main objective of present thesis is to expand current knowledge on the source of congruency effects in weakly spoken and recently trained second language (L2). It consists of six related experiments that measured interference produced on different levels of L2 proficiency. In all experiments within the present thesis, participants performed a bilingual Stroop task with intermixed native and second language words. Apart from measuring the magnitude of L2 interference, the present work investigated the source of this interference. That is, what produces the conflict between a colour word and an incongruent ink colour? By using the 2-to-1 mapping procedure (De Houwer, 2003), this interference has been decomposed into two subcomponents: stimulus and response conflict effects.

Experiment 1 conceptually replicated and extended the work of Schmidt and colleagues (2018), who investigated the congruency effects in a group of unbalanced Dutch-French bilinguals. They observed similarities in semantic and response processing of the native language and a moderately fluent second language. Experiment 1 aimed to test whether the same pattern occurs in a much less proficient L2. Therefore, we administered the same procedure used by Schmidt and colleagues (2018) on a group of French-English bilinguals with considerably lower L2 proficiency, evidenced by both subjective and objective measures. Thus, a crucial difference from the original study was considerably lower L2 competence. As in the original experiment, we observed significant stimulus and response conflict effects in response times for L2 words. Surprisingly, only the stimulus conflict effect, but not the response conflict effect, was significant for L1 (French) words. Although non-significant, L1 response conflict effect was numerically in the expected direction with faster responses on same response trials relative to different response trials. Furthermore, error data revealed the presence of response conflict effect for both L1 and L2 words. In contrast, stimulus conflict is rarely observed in the errors in the 2-to-1 mapping procedure and was unsurprisingly not observed for either language.

Most important for the present thesis are, of course, the L2 results. Our L2 results contrast the hypothesis that second language words produce stimulus conflict (Glaser & Glaser, 1989; MacKinnon et al., 1985) or response conflict (Klein, 1964; Sharma & McKenna, 1998) exclusively. As already discussed (see *Discussion* of Experiment 1), it seems that words from a weakly spoken L2 are potent enough to influence semantic processing and response selection similarly to first language words. Despite relatively low subjective and objective English competence in the sample, it is plausible that participants were familiar with English colour words. This limitation was thus considered in subsequent experiments. This potential caveat was addressed by training participants with completely unfamiliar L2 (Croatian) colour words.

**Word learning in L2**

A following series of experiments (Experiment 2-6) aimed to expand the present findings on a set of novel, recently trained L2 words. At the very beginning of these experiments, a short and simple learning procedure was administered. It remained unchanged for all experiments in the present thesis (Experiment 2-6). In the learning phase, four native language words from two different word types (i.e., French colour words in Experiment 2 and 3, English colour words in Experiment 4, and French colour associates in Experiment 5 and 6) were associated with their Croatian equivalents. A similar learning procedure was already reported in the literature (Altarriba & Basnight-Brown, 2012; Altarriba & Mathis, 1997). For instance, Altarriba and Mathis (1997) trained English monolinguals with a set of English-Spanish word pairs that were presented both visually on the screen and auditorily via headphones. During the learning phase, participants were passive and required to pay attention to word translations, similarly as in the present learning experiments (Experiment 2-6). In some other learning studies, novel L2 words were associated with their "descriptions" (Clay et al., 2007; Tamminen & Gaskell, 2013) or corresponding pictures (Clay et al., 2007; Dobel et al., 2010; Lotto & de Groot, 1998; Webber, 1978; Yu & Smith, 2007).

However, the initial learning phase requires higher level of engagement from participants. For instance, Geukes and colleagues (2015) administered a learning procedure in which word pairs consisted of one pseudoword and one colour word (e.g., "alep" and "blau"; i.e., German for blue). Participants were instructed to indicate whether the words belong together or not. Initially it was impossible for participants to tell whether a pair matched or not, but more frequent co-occurrence of certain "pseudoword-colour word" pairs in the learning phase helped discriminate matching from mismatching pairs.

To sum up, L2 word learning studies administered procedures that used different presentation forms (e.g., word translations, text, pictures, audio, etc.) and required different levels of subject engagement (e.g., active vs. passive). Taken together, initial learning aimed to familiarize participants with the to-be-memorized material that was further trained in subsequent phases of experiment.

**L2 word training**

The learning phase is usually followed by training phase that aims to strengthen a link between recently learned novel word forms and their lexical (i.e., L1 translations) and semantic (i.e., conceptual) representations. For instance, in the present series of experiments, novel Croatian (L2) words had to be matched with their L1 translations (Experiments 2-6), corresponding colour patch (Experiments 3 and 4), or corresponding image (Experiments 5 and 6).

These different types of training trials aimed to stress the semantic aspect of the L2 words. For the same purpose, Altarriba and Mathis (1997) adopted a set of quizzes in their training procedure. For instance, participants had to fill in the novel L2 word that corresponded either to a given L1 word or to a colour patch. The semantic aspect of L2 words was further emphasized by completing closed-ended L1 sentences with an L2 word, matching an emotion word with a particular L2 word, and matching an L2 word with an object that commonly appears in that colour. In the experiment conducted by Clay and colleagues (2007), training trials consisted of the presentation of a novel word acquired during the learning phase together with a description or a picture. Participants were required to decide whether the novel word matched with the presented description/picture. Although the training phase in these experiments were methodologically different, they were all focused on processing novel L2 words at a conceptual level.

The efficiency of novel L2 training has been evaluated through performance on the Stroop task (Altarriba & Basnight-Brown, 2012; Altarriba & Mathis, 1997; Clay et al., 2007; Geukes et al., 2015). Similarly, L2 word training from each experiment in the present thesis was followed by the Stroop task. Therefore, the features of the L2 word training will be discussed in the context of the Stroop interference effects produced by novel L2 words. The training performance was high in all experiments, suggesting that participants were able to memorize novel L2 forms and link them with corresponding lexical or semantic counterparts during relatively short word training. One important difference among the conducted experiments within the present thesis concerns a length and structure of training procedure, which will be discussed shortly.

For instance, Experiment 2 administered a relatively short (5-10 minutes long; 32 trials) and simple training procedure in which participants had to select a correct translation of a given target word (either L1 or L2) among four response alternatives. In this task, participants were encouraged to pay attention to novel L2 word forms (for shortcomings, see *Discussion* of Experiment 2). We observed a significant overall L2 congruency effect for response times, but its decomposition to stimulus and response conflict effects failed to reach significance. However, there was a hint toward L2 stimulus conflict for response times. Moreover, further data inspection revealed a significant L2 response conflict effect for the errors (although an overall Stroop interference was not significant). It seems plausible that the training was too short to sufficiently facilitate L2 semantic processing. However, the encouraging results showing a tendency toward L2 stimulus conflict in response times and a significant response conflict in errors motivated us to conduct a next experiment with additional training trials and slightly modified training phase.

The training procedure in Experiment 3 incorporated two new trial types that were supposed to strengthen the link between novel Croatian colour words and their corresponding concepts. The length of the training was extended relative to Experiment 2 (64 trials). Another adopted modification as compared to Experiment 2 concerned the addition of "catch" trials in a Stroop portion of the experiment (for more details, see the *Catch manipulation* section). Participants were instructed to withhold their response when one of the "catch" words were displayed on the screen. The purpose of this manipulation was to increase Stroop interference. Surprisingly, the pattern of results for L2 remained the same as in Experiment 2. That is, the response times for L2 words were in the expected direction with faster responses on identity relative to same response trials, although not significant. The L2 response conflict effect was evidenced, similarly as in Experiment 2. However, these two experiments did not allow us to address our research question, since no substantial L2 interference effects were produced. In other words, Experiments 2 and 3 essentially failed a manipulation check.

One potential explanation of these results concerns features of the administered training phase. In the following Experiment 4, we opted to simplify the structure of training trials by having only two response options (rather than four in Experiment 2 and Experiment 3) and multiplying the number of training trials (576 trials, 15-20 minutes long). Increasing the number of training trials should reinforce learning and the choice between two response alternatives was expected to facilitate responding (i.e., more trials in less time). Experiment 4 was run online on a sample of native English (L1) speakers. Experiment 4 yield striking results: a significant stimulus conflict effect in response times and response conflict effect in the errors were observed for recently trained Croatian words. This implies that novel foreign words are not only lexically connected to their native language equivalents (as suggested by the response conflict in the errors), but that there is also a very early semantic mediation (Altarriba & Mathis, 1997; Duyck & De Houwer, 2008). However, the presence of response conflict for novel L2 words argues against a model assuming exclusively semantic learning in early L2 learning. Interestingly, the response conflict effect was significantly reduced for L2 words as compared to L1 words for both response times and errors. Even though the primary goal of the present series of experiments was to determine the presence/absence of stimulus and response conflict for L2 words, these results suggest that response conflict is possibly reduced in early phases of L2 training.

These results partially support assumptions of the Kroll and Stewart (1994) model that suggest only the presence of response conflict effect for L2 words. According to their model, novel L2 words are translated directly to L1 equivalents, thus activating a corresponding response alternative. The presence of L2 response conflict effect in the errors is in line with this assumption. Furthermore, the model predicts no direct link with semantic representations should exist for novel L2 words at early stages of language acquisition. Our results showed that this semantic link (i.e., stimulus conflict) could be established for L2 words after relatively short training. Therefore, it is suggested that L2 words are potent enough to directly activate underlying semantic concept, without L1 mediation. Other results supported this notion of early semantic mediation in translation task. For instance, Duyck and Brysbaert (2004) investigated a semantic number magnitude effect in forward (L1-L2) and backward (L2-L1) translation. The standard finding is faster translation of numbers that represent small quantities (e.g., two) than larger quantities (e.g., eight). In a critical experiment, participants learned a set of pseudowords as translation equivalents of native Dutch number words. Interestingly, a large semantic effect of number magnitude was observed in both translation directions, even though the pseudowords had been acquired only recently. Their results suggest that novel word forms are mapped into semantics even in very early stages of L2 acquisition.

**Other theoretical accounts**

Apart from the Kroll and Stewart model (1994), there are some other models that could possibly account for the present findings. For instance, the *Bilingual Interactive Activation* model (van Heuven et al., 1998) assumes a non-selectivity of bilingual word recognition, that is, both languages are active during word processing. The focus is on the orthographic representation of words, assuming the existence of shared lexical systems in proficient bilinguals. This contrasts the assumption of the Kroll and Stewart (1994) model which clearly separates L1 and L2 lexicons. The language-nonselective access leads to the automatic co-activation of words from both lexicons when a printed word stimulus is presented. All activated orthographic representations (even those from the irrelevant language) compete for selection, which slows down word recognition. Top-down inhibitory control from language nodes limits the influence of this cross-language interference. The revised version of the original BIA model (i.e., BIA+) accounted for the role of semantics and phonology during lexical access.

However, it is important to emphasize the difference between models that focus word translation (e.g., the Kroll and Stewart model) and those that focus on word recognition (e.g., BIA, BIA+). Although the *Bilingual Interactive Activation* models account for a large range of bilingual language processes, they do not provide enough explanations for what happens during L2 development. In other words, neither the BIA model nor the BIA+ model make predictions for less proficient bilinguals and bilinguals at early stage of L2 acquisition, which is crucial for interpretation of present series of findings.

Another theoretical account that could possibly explain these findings is a model that combines the main features of the BIA model with a developmental aspect of second language acquisition, argued by Kroll and Stewart (1994). This model is known as the *developmental BIA model* (BIA-d; Grainger et al., 2010) and it predicts developmental changes in connectivity between L1 and L2 word form representations and semantics as function of L2 exposure. For instance, when a native English speaker that studies French encounters a novel L2 word "chaise" (French for chair), the L1 word forms "chain" and "chair" will be co-activated. L1 word forms are mutually inhibited but connected to their semantic representations. The exposure to a novel L2 word ("chaise") co-activates the equivalent L1 form ("chair") and the corresponding semantic representations. A direct link between the L2 word and semantics is further strengthened, which consequentially modifies the connection between L1 and L2 word forms. However, the limitation of the present model concerns the status of non-cognates in word recognition. That is, how do we recognize L2 words that do not overlap in pronunciation and spelling with L1 equivalents, as relevant for the present thesis?

Apart from the presented models, the Kroll and Stewart model (1994) remains the main theoretical viewpoint of the present research. Although it makes clear predictions about the nature and strength of lexicosemantic connections, it does not consider the content of lexical and semantic storages. The model that primary focuses on the organization of bilingual semantic representations is the *Distributed Feature Model*. This model clearly separates the lexical level that contains word forms, and the conceptual level that stores semantic representations. The written and spoken lexical representations of L1 and L2 are stored separately. However, the conceptual level consists of features shared between languages, that are activated by their corresponding words. For instance, the *Distributed Feature Model* assumes that concrete and abstract words differ in the degree of the semantic overlap across languages (de Groot, 1992). For instance, words in L1 and L2 that have a large overlap in meaning (i.e., concrete words, such as “door” in English and “porte” in French) share many conceptual features. In contrast, abstract words (e.g., “advice” in English and “conseil” in French) often have language-specific meanings and share a smaller proportion of the conceptual nodes between languages. Therefore, it seems plausible that novel L2 learning could partially depend on the type of to-be-learned word, that is, whether the L2 word is concrete or abstract word. In the present studies, only concrete words were used (i.e., colour words in Experiment 1-4 and colour associates in Experiment 5-6; for the corresponding discussion, see the following section). According to the assumptions of the *Distributed Feature Model*, these words could be easier to learn, since their translations across languages share large number of conceptual features (for further discussion, see *Implications for novel L2 word learning* section).

To sum up, a bilingual brain encounters multiple challenges during lexical access, that have been explained by different theoretical accounts. For instance, certain models focus on the word recognition and simultaneous activation of words in both lexicons. Some other models address competition between words with similar orthographies or semantic representations, as well as development of these connections.

## Colour associates

The interesting finding about very early semantic mediation (Altarriba & Mathis, 1997; Duyck & Brysbaert, 2004) for recently trained L2 colour words motivated us to further investigate the role of other word types (apart from colour words, as in the previous experiments) in foreign language acquisition. For instance, Klein (1964) was the first to measure the interference produced by different word types in a Stroop colour identification task (see *Introduction*). He observed a *semantic gradient,* that is, increased interference as a function of the relation between the word and colour. For instance, a colour associate ("sky") is a word strongly related to a colour ("blue") that produces more interference than colour-neutral words ("coat"). It has been argued that two stimulus dimensions (i.e., word meaning and ink colour) are associatively related, thus producing interference when a colour associate (e.g., "sky") is printed in incongruent colour (e.g., "green"). Other authors went further by investigating the source of this interference. They revealed that on incongruent trials (e.g., "sky" in green), the concurrent activation of the word and the target colour (i.e., "blue" and "green", respectively) produce a stimulus conflict effect (Schmidt & Cheesman, 2005). Colour associates are therefore used as an argument that the Stroop effect results, in part, from early, semantic processes. Following this logic, the associative strength between colour associates and corresponding colours could be possibly transferred to the novel words in the L2 word training procedure.

To test this notion, Experiment 5 and Experiment 6 trained participants with a set a Croatian colour associates paired with their French translations in a similar way as in Experiment 4. Important to note is that L2 colour associates have not been associated with colour words neither in learning nor in training phase. In other words, participants learned a novel L2 colour associate (e.g., "jagoda") paired with their L1 equivalent (e.g., "fraise", French for strawberry), but not the underlying colour concept (e.g., "red"). Therefore, Experiments 5 and 6 made use of another to-be-learned L2 word type and their corresponding L1 equivalents (i.e., not only colour words). In the Stroop portion of Experiments 5 and 6, apart from L1 and L2 colour associates, the Stroop task included L1 colour words closely related to colour associates. The presence of L1 colour words in the Stroop task should provide contextual information (see Geukes et al., 2015) and potentially increase the interference effect for colour associates. For instance, the presence of the colour words could help in activating the general semantic field of colour, which in turn may facilitate access to the specific meaning of given colour associates.

The overall performance in the training phase was good, assuming that participants were relatively successful in matching novel L2 colour associates with corresponding L1 equivalents or images. However, no substantial interference effects were observed in Experiment 5, neither for L1 nor L2 colour associates. It remained unclear whether this lack of interference effect is due to short training or a too small L2 associate effect. We opted to replicate Experiment 5 with addition of catch trial manipulation in order to boost the interference (Experiment 6). This manipulation resulted in significant stimulus and response conflict effects for L1 colour associates (Experiment 6), although previous evidence suggested the presence of exclusively stimulus conflict (Schmidt & Cheesman, 2005). Of particular interest were, of course, L2 associate effects where the results were more ambiguous. The main question was whether the novel L2 associates relate to their L1 translations through direct association and produce similar results as L1 associates. For instance, previous studies have shown that reinforced associations between two stimuli can trigger an interference in a Stroop task. MacLeod and Dunbar (1988) trained participants to name a colour when a particular shape was presented (e.g., when triangle was presented, participants had to say “yellow”, etc.). Extensive training reinforced these contingencies between shapes and colours. When performing a Stroop task, interference occurred when participants had to name the colour in which shapes were presented (MacLeod & Dunbar, 1988). Schmidt and colleagues (2007) used a Stroop-like procedure in which participants had to identify the ink colour of the printed colour-neutral word (e.g., “move”), which was mostly presented (e.g., 75% of times) in a particular colour (e.g., “blue”). Colour identification was facilitated when the word appeared in the associated colour (e.g., “move” in blue) relative to when it appeared in another colour (e.g., “move” in green; Schmidt et al., 2007). Recently, Liefooghe and colleagues (2020) trained their participants with a set of nonwords that were either: 1) directly reinforced for their connection with a colour word (i.e., *reinforced associate*: “plesk”) or 2) associated with a colour word through derivation (i.e., *derived associate*: “smelk”, for the detailed description of the training procedure, see *Introduction*). After the extensive training, both reinforced and derived associates produced a substantial Stroop effect (Experiment 2). Further decomposition of the interference effects produced by reinforced and derived associates revealed that these effects were driven by response conflict (Experiment 3-5), which supports the proposal of Kroll and Stewart (1994). As already discussed, their model argues that newly acquired L2 words are linked to their L1 translations only at the lexical level. Certain similarities between the series of studies by Liefooghe and colleagues (2020) and the present colour associate experiments could be pointed out. For instance, the novel L2 associates were directly reinforced through their association with L1 associates, similarly as *reinforced associates* in the study of Liefooghe and colleagues (2020). However, the notable difference is a lack of direct association with a particular colour word. That is, a novel Croatian word “jagoda” (strawberry) was associated with “fraise”, but not with “rouge” (French for red), which potentially limits the availability of the colour component during semantic activation. Moreover, responses in Experiments 5 and 6 were labelled with colour words, such as “rouge”, and not with associates (e.g., “fraise”). The substantial L2 associate effect did not occur, and thus, its decomposition to stimulus and response conflict effect seemed unreasonable. It is plausible that the L2 colour associates cannot be learned by using the administered learning and training procedure. For instance, the training was possibly too short, or types of training trials were not sufficiently adapted for the acquisition of L2 associates. Another explanation is that the L2 associate effect is simply too small, even though the effects were numerically larger in Experiment 6 than in Experiment 5.

**Magnitude of L1 and L2 interference effect**

Apart from the presence of interference effects for colour associates, the magnitude of its subcomponents could be discussed. For instance, a common finding is that L2 produces smaller interference effects than L1. This notion is supported in our Experiment 1 and 2 where we observed a sizeable difference in the magnitude of stimulus conflict effect across languages, as well as in Experiment 2-4 for the response conflict effect. However, Mägiste argued that the amount of conflict is proportional to mastery of L2. In a series of experiments, Mägiste (1984, 1985) compared the amount of interference in Swedish monolinguals and German-Swedish bilinguals. As expected, large interference was observed for monolinguals, but after mastering Swedish, Stroop interference became of comparable size in both languages for German-Swedish bilinguals. Note, however, that interference of equal magnitude for L1 and L2 occurred after years spent in an L2 environment (i.e., in Sweden).

The magnitude of the interference can also depend on other factors. One of them is a language mode, that refers to activation of the language or languages in the bilingual's mind. For instance, a native French speaker that has been talking, studying, or thinking in French for a while is in an "French mode" (Marian & Spivey, 2003). However, when a bilingual (e.g., the same native French speaker that also speaks English) switches between the two languages, both languages (i.e., French and English) become more salient, activating a "bilingual mode" (Grosjean, 1997). The language mode can be influenced by the experimental design, that is, when trials from both languages are randomized within the same experimental block, participants are engaged in a bilingual mode. This was the case in the present series of experiments, when both L1 and L2 distracters were presented interchangeably. This activation of the bilingual mode could possibly influence the magnitude of the two types of conflict, observed in the present studies. For instance, a comparable magnitude of stimulus conflict between L1 (French, English) and recently trained L2 (Croatian) words was observed in Experiment 3 and Experiment 4, respectively. Although our participants were not completely bilingual, they became experts on the set of studied foreign (Croatian) colour words (Altarriba & Mathis, 1997). The presentation of novel, recently trained L2 words together with their L1 equivalents with strong semantic representations, results in participants being in bilingual mode. Therefore, it seems plausible that the amount of L2 interference (or at least its components) could reach a level comparable to L1 after only short novel word training. Further investigations are needed to address this question.

**Effects of response language and response modality on interference**

However, the general consensus which supports the asymmetry between L1 and L2 congruency effects argues that it depends on the response language (Dyer, 1971; Preston & Lambert, 1969; Tzelgov et al., 1990). These are important caveats that should be considered in interpretation of the results. As already discussed, the present work made use of manual (i.e., keypress) responses, which is a requirement for applying a 2-to-1 mapping procedure. Possibly different patterns of results might occur with a verbal response modality. For instance, the magnitude of the Stroop interference is usually substantially larger when the subjects are required to identify the colour of the word verbally rather than manually (Augustinova et al., 2019; Redding & Gerjets, 1977; Sharma & McKenna, 1998). For instance, when a native French speaker is asked to produce verbal French responses (i.e., L1 in Experiment 1), the ink colour has to be named in French, regardless of the distracter language (i.e., either French as L1 or English as L2). This increases French colour-word interference and decreases English colour-word interference. The reverse would be the case with English vocal responses. In the present series of experiments, we used manual responses exclusively, which are not inherently compatible with either language, and therefore should not particularly contribute to the previously discussed asymmetry. Another argument for the chosen response modality is the fact that 2-to-1 mapping procedure (De Houwer, 2003) cannot be administered with verbal responses.

However, it is plausible that other differences between languages could have been explored if we had used a vocal (verbal) modality. For a native French speaker, L2 words like “green” or “zelena” might interfere less that L1 words (i.e., “vert”, in Experiment 1 and Experiments 2-3, respectively). This corresponds to the findings of Dyer (1971). For instance, when English monolinguals had to name the colour of Greek, Italian, and Spanish words, the amount of interference was proportional to similarity to their English equivalents. For instance, the Greek words for red, blue, and green share no similarity to any English words and were the least interfering. Spanish and Italian words were also far removed from their English translations and they induced only slightly more interference than Greek words (Dyer, 1971). It is reasonable therefore to expect that for a native French speaker, when asked to name the ink colour in French, a Croatian word “zelena” would be less interfering than its French equivalent “vert”.

However, the magnitude of interference produced in colour identification could depend on whether this colour belongs to the response set or not. This is known as the *response set membership* effect (Klein, 1964; Risko et al., 2006), which is a larger interference effects for words (i.e., distracters) that are potential targets (i.e., are assigned to one of the responses). This finding was confirmed with colour words (Klein, 1964; Sharma & McKenna, 1998) and colour associates (Risko et al., 2006). The latest experiment revealed that colour associates related to a colour in the response set interfere more than colour associates related to the colour out of the response set. However, colour associates unrelated to a colour in the response set produced more interference than colour-neutral words (Risko et al., 2006). In a French colour-naming task with four possible responses (e.g., “rouge”, “bleu”, “vert”, and “jaune”, French for red, blue, green, and yellow, respectively), the words “neige” and its English equivalent “snow” (related to white, which is not one of the potential responses) are expected to produce less interference than word “ciel” or English translation “sky” (related to blue, which is one of the potential responses). Following this logic, for a native French speaker who does not speak Croatian, word “neige” which is unrelated to a colour in the response set should produce more interference than its Croatian translation “snijeg” (even though this word is also colour-related). Similarly with the colour words, L2 words (e.g., “green” or “zelena”) are not potential targets and do not belong to the response set and therefore produce less interference than L1 words (“vert”). However, the source of this interference is not clear. As already mentioned, the 2-to-1 mapping manipulation does not lend itself well to verbal responses, but future work with alternative methodologies might address these possibilities.

This response modality effect could be possibly explained by different mechanisms that underlie manual and verbal responding (Kinoshita et al., 2017). For instance, verbal responses require colour naming, while manual responses involve colour identification. This difference between manual and verbal variants of the Stroop task emphasizes different mechanisms that lead to Stroop interference. This is in line with the important models of Stroop interference (Glaser & Glaser, 1989; Sharma & McKenna, 1998; Sugg & McDonald, 1994), which assume that the locus of this effect varies by response modality. Some other authors supporting the traditional response competition account (Cohen et al., 1990; Roelofs, 2003) of the Stroop task argued that the Stroop effect is produced by the reading task which remains unchanged regardless of the response modality. In other words, according to these accounts, the locus of the Stroop effect should be similar for both manual and verbal responding.

The following discussion will concern mostly the manual response modality, relevant for the present series of experiments. With manual responses, an irrelevant distracter activates one response option which contradicts another response option (e.g., a relevant dimension, target colour). However, manual responses could be labelled either with a word (i.e., “green”) or corresponding colour (i.e., green colour patch). For instance, translation accounts argue that the occurrence of Stroop interference effects depend on how the response buttons are labelled. According to this view, if the response buttons are labelled with words, interference should occur when participants are responding to the colour (i.e., standard Stroop task), but not to the word (i.e., reversed Stroop task). In contrast, if the response buttons are labelled with colour patches, interference should occur when participants are responding to the word, but not when responding to the ink colour. This account posit that an irrelevant distracter affects performance only when a translation is required to respond correctly to the target. Therefore, according to Sugg and McDonald (1994), only trials that require this translation module should produce interference, since it is assumed that colours and words are processed in separate systems, each operating in its own code (Glaser & Glaser, 1989). Response buttons in our experiments were always labelled with colour words, which can, according to translation accounts, explain the observed interference. Another possibility could be to label response keys with corresponding colour patches and observe the occurrence of L1/L2 interference effects and their subcomponents (e.g., stimulus and response conflict).

## Task conflict

As already discussed, the magnitude of Stroop interference was larger when verbal responses were used, relative to manual responses. This could be explained by the fact that Stroop interference observed with verbal responses arise from a significant contribution of task, stimulus, and response conflict effects. In contrast, only stimulus and response conflict are found to contribute to Stroop interference observed with manual responses (Augustinova et al., 2019). In other words, some authors argue for an additional contribution of task conflict in overall Stroop interference (Augustinova et al., 2018, 2019; Kinoshita et al., 2017). Task conflict represents slower colour identification of words and word-like stimuli relative to non-readable stimuli, such as letter strings, symbols or shapes. For instance, Kinoshita and colleagues (2017) compared five types of colour-neutral distracters, real words (e.g., “hat”), pseudowords (e.g., “hix”), consonant strings (e.g.,”htk” ), symbol strings (e.g., “%@$ ), and a row of Xs (e.g., “XXX”) with incongruent colour words (e.g., “green” in red) in manual and verbal variant of the Stroop task. Incongruent colour words produced robust interference as compared to colour-neutral stimuli in both variants of the task. With manual responses, all five types of colour-neutral distracters produced comparable response latencies. Interestingly, with verbal responses, relative to a row of Xs, words and pseudowords interfered equally and more than the consonant strings, which in turn interfered more than the symbols. This reflects a tendency to read a distracter word, even when this reading task is irrelevant for task performance and should be ignored (Kalanthroff et al., 2013; Monsell et al., 2001).

Task conflict is observed for all types of readable items, so it could possibly occur even for the recently trained foreign words (e.g., plava). Important to mention, in all of our experiments we used only words (i.e., no non-readable neutral stimuli) and participants were explicitly informed that they will be presented with a set of real Croatian colour words. This makes the task conflict less relevant for the present work, but it leaves room for further investigations. For instance, one possibility would be to investigate task conflict in an L2 learning paradigm. One way to achieve this would be to include word stimuli that are non-readable in the L1 language. Integrating word stimuli composed of characters specific for Croatian language (e.g., đ, ć, ž, nj), that are non-readable for French or English (L1) speakers (e.g., "žući", "crnji", etc.), would allow us to further examine a potential contribution of task conflict to Stroop interference.

**Catch manipulation**

As one potential caveat, the Stroop portion of certain experiments in this thesis (Experiments 3, 4, and 6), included two random Croatian words that served as catch trials. The aim was to increase the Stroop interference since participants were explicitly instructed to withhold responding to the two catch words. The meanings of these catch words were never given to the participants and were completely irrelevant for the conducted experiments. Therefore, participants based their response decision on word recognition (e.g., "Is this one of the two 'words' that I am not supposed to respond to?"). It is unlikely that these catch words were subjected to semantic analysis (e.g., "What is the meaning of this word and is it related to colours?"), simply because participants did not know their meanings. However, it is possible that this "catch trial" manipulation influences stimulus and response conflict differently, such that the exact size of each component might vary with versus without catch trials. On the other hand, it is unclear why the presence of catch trials did not influence response latencies similarly in all experiments that included this type of manipulation. For instance, in Experiment 4, catch trials (or another methodological feature) increased L2 Stroop interference, which was not the case in Experiment 3 and 6 where the L2 interference effect was non-significant. It is plausible that the L2 interference effect (Experiment 3) and L2 associate effect (Experiment 6) were too small due to insufficiently long or adapted training. As already discussed, the training procedure in Experiment 3 was relatively short (64 trials) and did not produce substantial effects. On the other hand, Experiment 6 focused on L2 colour associates learning, that is, words that are not colour words, but are strongly associated with colours. It remains possible that the administered word learning/training manipulation does not lend itself to this word type. For this reason, it did not result in a sufficiently large L2 interference effect. Related to this, a catch word manipulation may simply not be enough to "boost" the effect sufficiently.

**Cognate status**

The present results suggest that there is a certain similarity between L1 and L2 in the way they affect semantic and response processing (Dyer, 1971). This was observed for low proficient L2 words (i.e., English; see Experiment 1), but also for recently trained L2 words (i.e., Croatian, see Experiment 4). For instance, L2 colour words produced stimulus conflict (Experiment 4) or both stimulus and response conflict (Experiment 1), even though there was no phonological or orthographic overlap with L1 words. Indeed, as previously discussed (see Introduction), we opted to use foreign words that are dissimilar from their L1 translations (e.g., "vert" vs. "green" in Experiment 1; "blue" vs. "plava" in Experiment 2). Thus, L2 words (e.g., "green", "zelena") that are orthographically and phonologically dissimilar from their L1 equivalents are potent enough to interfere with semantic identification and response selection. This was confirmed with a sample of unbalanced French-English bilinguals (Experiment 1), that had low overall L2 proficiency despite the formal training (e.g., English classes in the school, etc.). A similar pattern of results (see *Experiment 4*) with L2 stimulus conflict in the response times and L2 response conflict in the errors was observed for a novel, obscure language that received only a short training (around 20 minutes). However, several other experiments failed to produce a sufficiently robust Stroop interference effect, which does not justify its decomposition to stimulus and response conflict. Considering the fact that cognates produce a stronger interference effect than non-cognates (Dyer, 1971; Preston &

Lambert, 1969), they are more suitable for investigating the source of the interference in recently trained L2 words.

### Implications for novel L2 word learning

The present work could be a good starting point in investigating the effectiveness of different types of L2 training and their effects on semantic and response processing. A large body of research has focused on investigating how specific mnemonic techniques can facilitate L2 acquisition. For instance, previous studies that are conceptually similar to learning experiments presented in this thesis (e.g., Experiment 2-6), used other presentation modalities to form connections between L1 and L2 translation equivalents. For instance, word pairs were linked by auditory (Altarriba & Basnight-Brown, 2012; Altarriba & Mathis, 1997) or pictorial (Altarriba & Knickerbocker, 2011; Lotto & de Groot, 1998) presentation. Other cognitive strategies include for example semantic mapping, in which semantically close words are presented together visually (Badr & Abu-Ayyash, 2019; Zahedi & Abdi, 2012), imagery by using key words (Atkinson & Raugh, 1974), and rhythmic speaking, singing, or music piece accompaniment (Degrave, 2019; Good et al., 2015; Ordin & Polyanskaya, 2015). Therefore, L2 learning procedures often engage different word presentation modalities. This is in line with the idea of the *Dual Coding Theory* which assumes that human memory comprises two separate but interacting systems: verbal memory (which deals with language; e.g., a novel L2 word) and image memory (specialized for the processing of nonverbal stimuli; e.g., image or sound). According to this theory, the chances that a memory will be retrieved are greater if the information in stored in two distinct systems rather than in just one (Clark & Paivio, 1991; Jared et al., 2013; Paivio et al., 1988).

Apart from initial presentation of to-be-learned L2 words, their associative strength with L1 counterparts was further reinforced in semantically elaborated tasks (Altarriba &

Basnight-Brown, 2012; Altarriba & Mathis, 1997). To sum up, there is a large variety of methods and techniques applied to L2 word acquisition that use different means to test their effectiveness. In the following sections, I discuss more on the possible adaptations of the learning/training procedures which are adopted within the scope of the present thesis (Experiments 2-6). In particular, I focus on the role of word type (i.e., concrete vs. abstract words) and L2 word organization (e.g., clustering), that could possibly facilitate semantic processing.

While the present series of experiments indicates that semantic and response influences are observed for some L2 words, one may wonder whether the present findings could be generalized to other word types, such as less frequent, emotional, or abstract words. For instance, Altarriba and Basnight-Brown (2012) trained English monolinguals on a set of concrete (e.g., jewel), emotion (e.g., angry), and abstract (e.g., virtue) Spanish words with a similar training procedure as the one used by Altarriba and Mathis (1997; see Introduction). After learning a set of novel words, participants performed a Stroop colour identification task. The analysis of response latencies revealed that emotion words were responded faster than concrete and abstract words. This contrasts with a typical emotional Stroop effect (i.e., slower response times for emotion words as compared to neutral words), reported for monolinguals. This lack of interference effect for novel L2 emotional stimuli can be explained by relatively weak and poorly developed semantic component of newly acquired emotional words.

This distinction between different word types in bilinguals is one of the core assumptions of the *Distributed Feature Model* (see the *Other theoretical accounts* section), which states that concrete and abstract words differ in the degree of the semantic overlap across languages (de Groot, 1992). As already discussed, according to this model, concrete words in L1 and L2 have a large overlap in meaning and they share many conceptual features. In contrast, abstract words often have language-specific meanings and share a smaller proportion of the conceptual nodes between languages. Words with a higher level of abstraction (i.e., abstract or emotion words) are deeply encoded in L1 (Pavlenko, 2009), due to strong connection between those words and the context in which they are learned. Thus, it seems plausible that the learning of certain word types in L2 would be facilitated if they are presented within a context (e.g., written, spoken, or visual). Acquiring novel words in the corresponding context (e.g., that complements the valence of L2 emotion word) could help building a “cue environment” that could facilitate later retrieval or recognition of that word (Altarriba & Basnight-Brown, 2012; de Groot, 1992; Schwanenflugel et al., 1992).

Second language learners are often presented with novel words organized in sets of semantically similar words, or “semantic clusters” (e.g., “eye”, “nose”, “ear”, “mouth”, or in colour terms: “red”, “blue”, “green”, “yellow”). However, based on the interference theory, it was assumed that this grouping of to-be-learned items impede rather than facilitate L2 vocabulary learning. For instance, in one experiment, learning of novel L2 words was carried out across different modalities (i.e., oral and written), and was further tested on recall and recognition tasks. The results showed that the sets of artificial words paired with semantically related English words (e.g., “apple”, “pear”, “plum”) were more difficult to learn relative to artificial words paired with unrelated English words (e.g., “paint”, “uncle”, “ice”). This finding implies that semantic clustering inhibits the learning of novel L2 vocabulary (Tinkham, 1997). Semantic clustering was used in the present series of studies with the set of semantically related French colour words (i.e., “vert”, “bleu”, “gris”, “rouge”, in English “green”, “blue”, “grey”, “red”, respectively; see Experiment 2 and 3), in which L1 words were paired with to-be-learned Croatian words.

However, it is possible that there is a more optimal word clustering alternative, so called thematic clustering. Thematic clusters consist of words that are all closely related with a common thematic concept. For instance, a set of to-be-learned L2 words (e.g., “zelena”, “trava”, “bor”, “maslina”) could be paired with its equivalents from L1 thematic cluster (e.g., “vert”, “herbe”, “pin”, “olive”; in English “green”, “grass”, “pine”, “olive”, respectively). All the words within the cluster are closely related with a common thematic concept, which could be a more efficient alternative that facilitates L2 vocabulary learning (Tinkham, 1997).

To sum up, theoretical and empirical evidence suggests that both the learning method and word type might play important roles in establishing connections between novel L2 words and their semantic representations. However, the present series of experiments focused on standard colour-word Stroop procedure that is inherently limited to the use of colour-related stimuli. However, current findings could be expanded with “word-word” Stroop task variants that can be used with any word type (Glaser & Glaser, 1982, 1989; Schmidt et al., 2013). For instance, the interference effect has been observed in a word-word variant of the Stroop task when both the distracter and the target are words. That is, responses are typically slower when the target word (e.g., “green”) is primed with an incongruent colour (e.g., “red”) than with a neutral word (e.g., “move”) or a congruent word (e.g., “green”). Administrating the similar procedure, a recently trained L2 word (e.g., “zelena”, Croatian for green) could serve as a prime in target identification task. With extensive practice, L2 primes could influence the target identification speed. In other words, it seems plausible that the magnitude of interference produced by novel obscure L2 words would vary as a function of L2 practice. Future investigations are needed to shed light on this issue.

**Conclusions**

This thesis provides an extensive overview on the source of interference effects in weakly spoken and recently learned second language. The six experiments presented here revealed that novel L2 words are acquired relatively rapidly and associated with both L1 translations and semantic representations. Thus, there is a certain similarity between L1 and L2 words in the way they influence semantic processing and response selection, which contrasts with the proposals of Kroll and Stewart (1994). However, the development of these lexical and conceptual connections at least partially depends on the training characteristics (e.g., length, number, and type of training trials, etc.) and to-be-learned word type (e.g., colour words, colour associates, etc.). The present results contributed to better understanding of cognitive processes involved in very early language learning.

## References


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

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

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

Atkinson, R. C., & Raugh, M. R. (1974). An application of the mnemonic keyword method to the acquisition of a Russian vocabulary. *Psychology and Education Series, Technical Report No. 237*, 40.

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

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

Augustinova, M., Silvert, L., 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*(2), 543–549. https://doi.org/10.3758/s13423-014-0697-z

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

Badr, H. M., & Abu-Ayyash, E. A. S. (2019). Semantic mapping or rote memorisation: Which strategy is more effective for students' acquisition and memorization of L2 vocabulary? *Journal of Education and Learning*, *8*(3), 17.

Barcroft, J. (2003). Effects of questions about word meaning during L2 Spanish lexical learning. *The Modern Language Journal*, *87*(4), 546–561.

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

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

Bowers, J. M., & Kennison, S. M. (2011). The role of age of acquisition in bilingual word translation: Evidence from Spanish-English bilinguals. *Journal of Psycholinguistic Research*, *40*(4), 275–289. https://doi.org/10.1007/s10936-011-9169-z

Brown, M., & Besner, D. (2001). On a variant of Stroop's paradigm: Which cognitions press your buttons? *Memory & Cognition*, *29*(6), 903–904.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind*, *41*, 63–65.

Chen, H.-C., & Leung, Y.-S. (1989). Patterns of lexical processing in a nonnative language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 316–325.

Chen, H.-C., & Ng, M.-L. (1989). Semantic facilitation and translation priming effects in Chinese-English bilinguals. *Memory & Cognition*, *17*(4), 454–462. https://doi.org/10.3758/BF03202618

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review*, *3*(3), 149–210. https://doi.org/10.1007/BF01320076

Clay, F., Bowers, J. S., Davis, C. J., & Hanley, D. A. (2007). Teaching adults new words: The role of practice and consolidation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 970–976. https://doi.org/10.1037/0278-7393.33.5.970

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing account of the Stroop effect. *Psychological Review*, *97*(3), 332–361. https://doi.org/10.1037/0033-295X.97.3.332

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, *104*(3), 268–294. https://doi.org/10.1037/0096-3445.104.3.268

Crump, M. J. C., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLoS ONE*, *8*(3), e57410. https://doi.org/10.1371/journal.pone.0057410

de Bot, K. (2004). The multilingual lexicon: Modelling selection and control. *International Journal of Multilingualism*, *1*(1), 17–32. https://doi.org/10.1080/14790710408668176

de Groot, A. M. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 1001–1018. https://doi.org/10.1037/0278-7393.18.5.1001 de Groot, A. M., Dannenburg, L., & van Hell, J. G. (1994). Forward and backward word translation by bilinguals. *Journal of Memory and Language*, 600–629.

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

Degrave, P. (2019). Music in the foreign language classroom: How and why? *Journal of Language Teaching and Research*, *10*(3), 412. https://doi.org/10.17507/jltr.1003.02

Dobel, C., Junghöfer, M., Breitenstein, C., Klauke, B., Knecht, S., Pantev, C., & Zwitserlood, P. (2010). New names for known things: On the association of novel word forms with existing semantic information. *Journal of Cognitive Neuroscience*, *22*(6), 1251–1261. https://doi.org/10.1162/jocn.2009.21297

Duyck, W., & Brysbaert, M. (2004). Forward and backward number translation requires conceptual mediation in both balanced and unbalanced bilinguals. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(5), 889–906. https://doi.org/10.1037/0096-1523.30.5.889

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

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

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

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

Fraisse, P. (1969). Why is naming longer than reading? *Acta Psychologica: Attention and Performance*, 96–103.

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

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

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

Good, A. J., Russo, F. A., & Sullivan, J. (2015). The efficacy of singing in foreign-language learning. *Psychology of Music*, *43*(5), 627–640. https://doi.org/10.1177/0305735614528833

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

Grosjean, F. (1997). The bilingual individual. *Interpreting*, *2*(1/2), 163–187.

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

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

Heredia, R. R., & Cieślicka, A. B. (Eds.). (2015). In *Bilingual figurative language processing*. Cambridge University Press. https://doi.org/10.1017/CBO9781139342100

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology*, *4*(1), 11–26. https://doi.org/10.1080/17470215208416600

Hummel, K. M. (2010). Translation and short-term L2 vocabulary retention: Hindrance or help? *Language Teaching Research*, *14*(1), 61–74. https://doi.org/10.1177/1362168809346497

Jared, D., Pei Yun Poh, R., & Paivio, A. (2013). L1 and L2 picture naming in Mandarin–English bilinguals: A test of Bilingual Dual Coding Theory. *Bilingualism: Language and Cognition*, *16*(2), 383–396. https://doi.org/10.1017/S1366728912000685

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

Keatley, C. W., Spinks, J. A., & De Gelder, B. (1994). Asymmetrical cross-language priming effects. *Memory & Cognition*, *22*(1), 70–84. https://doi.org/10.3758/BF03202763

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

Kirsner, K., Brown, H. L., Abrol, S., Chadha, N. K., & Sharma, N. K. (1980). Bilingualism and lexical representation. *Quarterly Journal of Experimental Psychology*, *32*(4), 585–594. https://doi.org/10.1080/14640748008401847

Kirsner, K., Smith, M. C., Lockhart, R. S., King, M. L., & Jain, M. (1984). The bilingual lexicon: Language-specific units in an integrated network. *Journal of Verbal Learning* *and Verbal Behavior*, *23*(4), 519–539. https://doi.org/10.1016/S0022-5371(84)90336-0

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.

La Heij, W., Hooglander, A., Kerling, R., & van der Velden, E. (1996). Nonverbal context effects in forward and backward word translation: Evidence for concept mediation. *Journal of Memory and Language*, *35*(5), 648–665. https://doi.org/10.1006/jmla.1996.0034

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

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

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

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

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

MacLeod, C. M. (1998). Training on integrated versus separated Stroop tasks: The progression of interference and facilitation. *Memory & Cognition*, *26*(2), 201–211. https://doi.org/10.3758/BF03201133

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 126–135. https://doi.org/10.1037/0278-7393.14.1.126

Mägiste, E. (1984). Learning a third language. *Journal of Multilingual and Multicultural Development*, *5*(5), 415–421. https://doi.org/10.1080/01434632.1984.9994170

Mägiste, E. (1985). Development of intra- and interlingual interference in bilinguals. *Journal of Psycholinguistic Research*, *14*(2), 137–154. https://doi.org/10.1007/BF01067626

Manwell, L. A., Roberts, M. A., & Besner, D. (2004). Single letter coloring and spatial cuing eliminates a semantic contribution to the Stroop effect. *Psychonomic Bulletin & Review*, *11*(3), 458–462. https://doi.org/10.3758/BF03196595

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)

Marian, V., & Spivey, M. (2003). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics, 24*, 173–193.

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

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. https://doi.org/10.1037/h0031564

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

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326. https://doi.org/10.1037/0033-2909.132.2.297

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

Nikolov, M., & Mihaljević Djigunović, J. (2011). All shades of every color: An overview of early teaching and learning of foreign languages. *Annual Review of Applied Linguistics*, *31*, 95–119. https://doi.org/10.1017/S0267190511000183

Ordin, M., & Polyanskaya, L. (2015). Acquisition of speech rhythm in a second language by learners with rhythmically different native languages. *The Journal of the Acoustical Society of America*, *138*(2), 533–544. https://doi.org/10.1121/1.4923359

Paivio, A., Clark, J. M., & Lambert, W. E. (1988). Bilingual dual-coding theory and semantic repetition effects on recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 163–172.

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

Posner, M. I., & Presti, D. E. (1987). Selective attention and cognitive control. *Trends in Neurosciences*, *10*(1), 13–17.

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.

Potter, M. C., So, K.-F., Eckardt, B. V., & Feldman, L. B. (1984). Lexical and conceptual representation in beginning and proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior*, *23*(1), 23–38. https://doi.org/10.1016/S0022-5371(84)90489-4

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

Proctor, R. W. (1978). Sources of color-word interference in the Stroop color-naming task. *Perception & Psychophysics*, *23*(5), 413–419. https://doi.org/10.3758/BF03204145

Redding, G. M., & Gerjets, D. A. (1977). Stroop effect: Interference and facilitation with verbal and manual responses. *Perceptual and Motor Skills*, *45*(1), 11–17. https://doi.org/10.2466/pms.1977.45.1.11

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

Risko, E. F., Stolz, J. A., & Besner, D. (2005). Basic processes in reading: Is visual word recognition obligatory? *Psychonomic Bulletin & Review*, *12*(1), 119–124. https://doi.org/10.3758/BF03196356

Roelofs, A. (2003). Goal-referenced selection of verbal action: Modelling attentional control in the Stroop Task. *Psychological Review*, *110*(1), 88–125.

Scarborough, D. L., Gerard, L., & Cortese, C. (1984). Independence of lexical access in bilingual word recognition. *Journal of Verbal Learning and Verbal Behavior*, *23*(1), 84–99. https://doi.org/10.1016/S0022-5371(84)90519-X

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

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

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

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

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

Seymour, P. H. K. (1974). Stroop interference with response, comparison, and encoding stages in a sentence-picture comparison task. *Memory & Cognition*, *2*(1), 19–26. https://doi.org/10.3758/BF03197486

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

Shelton, J. R., & Martin, R. C. (1992). How semantic is automatic semantic priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(6), 1191–1210. https://doi.org/10.1037/0278-7393.18.6.1191

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

Smith, J. K. (1982). Converging on correct answers: A peculiarity of multiple-choice items. *Journal of Educational Measurement*, *19*(3), 211–220. https://doi.org/10.1111/j.1745-3984.1982.tb00129.x

Snodgrass, J. G. (1984). Concepts and their surface representations. *Journal of Verbal Learning and Verbal Behavior*, *23*(1), 3–22. https://doi.org/10.1016/S0022-5371(84)90479-1

Stirling, N. (1979). Stroop interference: An input and an output phenomenon. *Quarterly Journal of Experimental Psychology*, *31*(1), 121–132. https://doi.org/10.1080/14640747908400712

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.

Sugg, M. J., & McDonald, J. E. (1994). Time course of inhibition in color-response and word-response versions of the Stroop task. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 647–675.

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

Tamminen, J., & Gaskell, M. G. (2013). Novel word integration in the mental lexicon: Evidence from unmasked and masked semantic priming. *Quarterly Journal of Experimental Psychology*, *66*(5), 1001–1025. https://doi.org/10.1080/17470218.2012.724694

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

Tanaka, J., Weiskopf, D., & Williams, P. (2001). The role of color in high-level vision. *Trends in Cognitive Sciences*, *5*(5), 211–215. https://doi.org/10.1016/S1364-6613(00)01626-0

Tinkham, T. (1997). The effects of semantic and thematic clustering on the learning of second language vocabulary. *Second Language Research*, *13*(2), 138–163. https://doi.org/10.1191/026765897672376469

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*(5), 352–373. https://doi.org/10.1037/h0020071

Tzelgov, J., & Eben-Ezra, S. (1992). Components of the between-language semantic priming effect. *European Journal of Cognitive Psychology*, *4*(4), 253–272.

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

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

Webber, N. E. (1978). Pictures and words as stimuli in learning foreign language responses. *The Journal of Psychology: Interdisciplinary and Applied*, *98*, 57–63.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *18*(5), 414–420. https://doi.org/10.1111/j.1467-9280.2007.01915.x

Zahedi, Y., & Abdi, M. (2012). The effect of semantic mapping strategy on EFL learners' vocabulary learning. *Procedia - Social and Behavioral Sciences*, *69*, 2273–2280. https://doi.org/10.1016/j.sbspro.2012.12.198

Zhao, J., Liang, W.-K., Juan, C.-H., Wang, L., Wang, S., & Zhu, Z. (2015). Dissociated stimulus and response conflict effect in the Stroop task: Evidence from evoked brain potentials and brain oscillations. *Biological Psychology*, *104*, 130–138. https://doi.org/10.1016/j.biopsycho.2014.12.001