Cognition 165 (2017) 10–25

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Original Articles

Is susceptibility to cross-language interference domain specific?

Anat Prior *, Tamar Degani, Sehrab Awawdy, Rana Yassin, Nachshon Korem

University of Haifa, Israel

ARTICLE INFO

Article history:
Received 29 October 2015
Revised 6 April 2017
Accepted 20 April 2017
Available online 28 April 2017

Keywords:
Bilingualism
Interference
Cognitive control

ABSTRACT

The ability to overcome interference from the first-language (L1) is a source of variability in second language (L2) achievement, which has to date been explored mainly in same-script bilinguals. Such interference management, and bilingual language control more generally, have recently been linked to domain general executive functions (EF). In the current study, we examined L2 proficiency and executive functions as possible predictors of susceptibility to L1 interference during L2 processing, in bilinguals whose languages do not share an orthographic system. Seventy Arabic-Hebrew bilingual university students performed two tasks indexing cross-language interference (from L1 to L2). Lexical interference was assessed using a cross-modal semantic similarity judgment task in Hebrew, with false-cognates as critical items. Syntactic interference was assessed using a self-paced reading paradigm and grammaticality judgments on Hebrew sentences whose syntactic structures differed from those of Arabic. EFs were examined using spatial and numerical Stroop tasks, to index inhibitory control, and a task switching paradigm, to index shifting abilities. We found significant L1 interference across the lexical and syntactic domains, even in proficient different-script bilinguals. However, these interference effects were not correlated, and neither type of interference was related to domain general EF abilities. Finally, offline susceptibility to syntactic interference, but not lexical interference, was reduced with greater L2 proficiency. These results suggest at least partially independent mechanisms for managing interference in the two language domains, and raise questions regarding the degree to which domain general control abilities are recruited for managing L1 interference.

1. Introduction

The two languages of bilingual speakers are simultaneously active, requiring bilinguals to continuously manage potential interference from the non-target language (Kroll, Bobb, & Hoshino, 2014). Such interference is pervasive, and has been documented in language production (Hermans, Bongaerts, De Bot, & Schreuder, 1998) and comprehension (Dijkstra & Van Heuven, 2002) for single words and in sentence context (Libben & Titone, 2009). Critically, most previous research examined bilinguals who use two languages that share the Roman alphabet, such as English-Dutch or Catalan-Spanish. The extent to which such interference is characteristic of bilinguals who speak languages which differ in orthography is less well established (Giezen, Blumenfeld, Shook, Marian, & Emmorey, 2015; Hoshino & Kroll, 2008; Morford, Kroll, Piñar, & Wilkinson, 2014; Sunderman & Priya, 2012). Thus, the first goal of the present study is to examine to what degree cross-language interference is evident in proficient bilinguals who use different-script languages (Arabic-Hebrew).

Further, interference from the first-language (L1) while processing the second-language (L2) is evident in different aspects of language processing, including accent, lexicon and grammar (MacWhinney, 2005), but these domains have mostly been investigated independently in the past. In the current study we examine the ability of individual bilinguals to manage interference in both lexical and grammatical processing. Our second goal is therefore to adopt an individual differences approach to probe to what extent interference management is a generalized ability of the linguistic system. Namely, is interference management in the lexical domain tied to interference management in the grammatical domain? Further, we test whether greater L2 proficiency is associated with improved interference management across these two language domains.

Finally, cross-language interference management has recently been linked to domain-general executive function abilities (Blumenfeld & Marian, 2013; Pivneva, Mercier, & Titone, 2014). Our third goal, therefore, is to examine whether individual differ-

* Corresponding author at: Faculty of Education, University of Haifa, Mount Carmel, Haifa, Israel.

E-mail address: aprior@edu.haifa.ac.il (A. Prior).

http://dx.doi.org/10.1016/j.cognition.2017.04.006
0010-0277/

ences in interference management are linked to individual differences in domain general control mechanisms.

1.1. Cross-language interference

1.1.1. Lexical domain

To examine cross-language interference, studies typically capitalize on words that might create competition, such as false-cognates (also called interlingual homographs or homophones) which overlap in form but not in meaning across languages (Dijkstra, 2005). For same-script bilinguals, false-cognates typically share both orthographic and phonological form. For different-script bilinguals, in contrast, only the phonological form is shared across the two languages, arguably creating less potential for cross-language interference. Most previous research examining cross-language interference has focused on same-script bilinguals, and provided evidence of activation of the non-target meaning of false-cognates (for a review, see Degani & Tokowicz, 2010). However, much less is known regarding cross-language interference in processing false-cognates in the two languages of different-script bilinguals.

In same-script bilinguals, presenting false-cognates in writing allows for bottom-up meaning activation in both languages (but see e.g., Friesen & Jared, 2012). To create a similar situation for different-script bilinguals, in the current study false-cognates were presented aurally, thus providing bottom-up activation for both languages. We probed cross-language interference utilizing a semantic decision task, in which the activation of the non-target language (L1) unequivocally interferes with task performance (in the L2). In a recent study, Friesen and Jared (2012) showed that the meaning of interlingual homophones (overlapping in phonology and not orthography) in the non-target language interfered with bilinguals' semantic category decision. Specifically, French-English bilinguals were more likely to erroneously verify category membership of an interlingual homophone visually presented in English (shoe as a vegetable) when the French meaning of the word belonged to the probed category ("chou", which shares pronunciation with "shoe", means cabbage in French). We are unaware of parallel research in different-script bilinguals demonstrating semantic interference as a result of meaning activation of false-cognates in the non-target language (for phonological effects in masked priming in the absence of orthographic overlap see Kim & Davis, 2003, for Korean-English; Dimitropoulou, Duñabeitia, & Carreiras, 2011, for Greek and Spanish; Nakayama, Verdonschot, Sears, & Lupker, 2014, for Japanese English).

There is indirect evidence supporting the notion of language non-selective semantic activation via phonology. Lagrou, Hartsuiker, and Duyck (2013) demonstrated that aurally presented interlingual homophones activated their meanings in both languages of Dutch-English bilinguals. This cross-language activation was modulated but not eliminated by semantic constraint and speaker accent. In different-script bilinguals, Marian and Spivey (2003b) demonstrated that in a single language context, phonological input activated concepts across the two languages of Russian-English bilinguals, using a visual world paradigm (see also Marian & Spivey, 2003a). Thus, phonological input in one language likely leads to non-selective activation of lexical and semantic information in both languages of different-script bilinguals.

In the current study we examine this issue using a cross-modal semantic decision task on L2 word pairs. In critical trials, the aurally presented first word is a false-cognate between L1 and L2 of different-script Arabic-Hebrew bilinguals, and the second word is related to the meaning of the false-cognate in the non-target language. Thus, we examine whether the L1 meaning of a false-cognate presented aurally in the L2 can interfere with semantic decisions in the L2.

1.1.2. Grammatical domain

Cross-language interference in the grammatical domain has been investigated by examining how bilinguals process structures that are similar or different across the two languages. Interference is presumed when cross-language differences hinder processing (e.g. Nitschke, Kidd, & Serratrice, 2010; for a review see Kotz, 2010). For example, in an ERP grammaticality judgment task, Tokowicz and MacWhinney (2005) showed that L2 learners of Spanish were more error prone on grammatical structures not shared with their L1 English (unique to the L2) than on structures that exist in both languages. In the same study, ERP data showed greater sensitivity to violations that occurred in structures that were similar across the L1 and the L2 than in structures that differed cross-linguistically. Sabourin and Stowe (2008) investigated the sensitivity of L2 learners of Dutch to grammatical gender violations in Dutch, using ERPs. They found that L1 speakers of German, which has a similar grammatical gender system to Dutch, processed violations in a manner similar to that of native Dutch speakers, whereas L1 speakers of Romance languages, which differ in the grammatical gender system, did not. Further, Dussias (2003) found that English-Spanish bilinguals demonstrated syntactic parsing preferences in the L2 Spanish that were similar to the preferences prevalent in English, the L1, supporting the notion of transfer in the syntactic domain (MacWhinney, 2005). Similarly, Roberts, Gullberg, and Indefrey (2008) found evidence supporting the influence of L1 Turkish on pronoun resolution in L2 Dutch.

Tokowicz and Warren (2010) examined a similar question using a self-paced reading task. They found that English speaking beginning adult L2 learners of Spanish showed online sensitivity to grammatical violations in the L2 only in structures that are similar to those of the L1, but not in L2-unique structures. In contrast, in a second sentence reading study, Tuninetti, Warren, and Tokowicz (2014) reported no evidence for cross-language influence because participants' performance in English (L2) was not influenced by the status of the violation in their L1 (Arabic or Chinese). Notably, participants in the Tuninetti et al. (2014) study were more advanced L2 learners than participants in Tokowicz and Warren (2010), and also had L1s that differed in script from the tested L2. Thus, proficiency and/or script overlap might have led to the observed differences in performance. In addition, the later study investigated a highly salient grammatical structure (word order), and the L2 participants were very accurate in identifying violations, possibly masking L1 influences on performance.

In the current study we employ a similar paradigm to investigate cross-language influence in grammatical processing, using a self-paced reading task, with proficient bilinguals of different-script languages. Participants read sentences in the L2, half of which included grammatical structures that are similar across L1 and L2, and others with grammatical structures that differ across the two languages. Similar-structure and different-structure sentences could be either grammatically correct or include a grammatical violation. We employed a wide variety of grammatical violations in the L2, not all of which are highly salient, because less salient structures might be more sensitive to interference from the L1, especially in proficient bilinguals.

As detailed above, cross-language interference has been less investigated in different-script bilinguals. Moreover, findings from same-script bilinguals might not necessarily generalize to different-script bilinguals for two reasons. First, differences in script could theoretically reduce cross-language activation when processing written words by cuing bilinguals to the target language (Gollan, Forster, & Frost, 1997, but see Thierry & Wu, 2007). Second, even when processing spoken language, cross-language activation may vary with script overlap because the language system of different-script bilinguals may have evolved slightly differently (Sunderman & Priya, 2012), with greater separation and less overlap in activation between the two languages (but see, Hoshino & Kroll, 2008). Thus, the first goal of the present study is to extend our understanding of cross-language interference in proficient different-script bilinguals.

In addition, lexical and grammatical interference have to date been investigated as two separate phenomena. Based on current findings, we do not know whether interference management in the lexical domain is linked to interference management in the grammatical domain. In the current study, we measured the interference management abilities of a single group of participants across both domains, allowing us to test the possible link between them.

1.2. Individual differences in interference management

There is great individual variability in many aspects of second-language acquisition and processing (Dörnyei & Skehan, 2003; Roberts, 2012). It is therefore logical to assume that individuals could differ in their ability to manage cross-language interference. Interestingly, such variability may be linked to domain-specific (linguistic) and/or domain-general (cognitive control) mechanisms. In the current study we specifically explore whether individual differences in interference management are linked to individual differences in L2 proficiency and/or to individual differences in cognitive control.

1.2.1. Proficiency

Proficiency in the L2 may modulate L1 cross-language interference in more than one way. First, increased proficiency may change the balance of baseline activation of the two languages, such that L2 representations are more accessible and/or L1 representations become less accessible (Guo, Liu, Misra, & Kroll, 2011; Linck, Kroll, & Sunderman, 2009; Van Hell & Tanner, 2012). Second, more proficient L2 users may have gained improved abilities to overcome interfering information from the L1 when it is activated. These mechanisms are not mutually exclusive and could be operating in concert. If this is the case, then bilinguals who are more balanced or who are more proficient in the L2 should show reduced effects of L1 interference when processing L2. In the current study we examine whether increased L2 proficiency leads to improved L1 interference management in the lexical domain and in the grammatical domain.

In one previous study, Chambers and Cooke (2009) examined the impact of L2 proficiency and sentence context on activation of cross-language competitors in a visual world paradigm. The results showed that native English speakers who rated themselves as more proficient in L2 (French) were as likely to consider interlingual competitors as were less proficient participants. Thus, these findings do not support the notion of a link between proficiency and cross-language activation. However, the number of items per condition was very small (n = 3) as was the number of participants (n = 20), making it more difficult to find a link between L2 proficiency and cross-language interference management. Thus, one goal of the current research is to investigate this issue more systematically.

1.2.2. Cognitive control

Bilingual language processing, which inherently entails competition between the two languages, has been recently linked with domain general executive abilities (Kroll et al., 2014). More specifically, domain-general cognitive control has been linked to bilinguals' ability to overcome cross-language activation and interference. For example, Linck, Hoshino, and Kroll (2008) found that bilinguals who had larger working memory capacity showed smaller cognate facilitation effects in a sentence context, suggesting an improved ability to limit activation to the target language. The same study also found that bilinguals who had better inhibitory control (as indexed by a Simon task) had smaller cognate facilitation effects in a picture naming paradigm, again suggesting that they were less influenced by cross-language activation.

Further, when switching languages, bilinguals need to overcome activation of the previously used language in order to produce the target language. Bilinguals' facility in such language switching has been linked to their general ability to switch between non-linguistic tasks (Prior & Gollan, 2011; see also, Festman, Rodriguez-Fornells, & Münte, 2010), and to their inhibitory control abilities (Linck, Schwieter, & Sanderman, 2012).

Finally, using a visual world paradigm, Blumenfeld and Marian (2013) recently showed that English-Spanish bilinguals with stronger inhibitory control, measured by a Stroop task, were *more* likely to activate L2 cross-language competitors but were also more efficient in overcoming such activation. This relation held for both higher and lower proficiency bilinguals. Mercier, Pivneva, and Titone (2014) found a somewhat different pattern, where bilinguals with higher cognitive inhibitory control were *less* likely to activate cross-language competitors (see also Pivneva et al., 2014, for evidence from an L2 sentence reading task). Critically, both studies found a significant link between domain-general inhibitory control and the dynamics of cross-language activation and interference. Notably, we are currently unaware of any studies that have directly investigated this link in the grammatical domain.

Thus, our final goal in the current study is to examine the possible impact of L2 proficiency and domain-general executive control on bilinguals' ability to overcome L1 interference when processing the L2. As stated above, these two factors are not mutually exclusive, and might concurrently influence performance.

To achieve the outlined goals, in the current study a group of proficient Arabic-Hebrew bilinguals performed a battery of executive control tasks, measures of L2 proficiency, and two experimental language tasks in Hebrew, the L2. Executive control was examined using a spatial Simon paradigm (based on Bialystok, Craik, & Luk, 2008; Mor, Yitzhaki-Amsalem, & Prior, 2014), a numeric Stroop paradigm (based on Hernandez, Costa, Fuentes, Vivas, & Sebastian-Galles, 2010; Mor et al., 2014), and a task switching paradigm (Prior & MacWhinney, 2010). Measures of L2 proficiency included both subjective self-ratings and objective measures of Hebrew proficiency in both the lexical and the grammatical domains (see Method section below for details).

The experimental language tasks were specifically tailored to elicit cross-language interference from Arabic, the L1, and to probe participants' ability to overcome such interference when processing Hebrew, the L2. In the lexico-semantic domain, participants performed a cross-modal semantic decision task on pairs of words presented in the L2. Critical trials included false-cognates as the first word and words that were related to the L1 meaning of the false-cognate as targets. In the grammatical domain, participants read L2 sentences in a self-paced reading paradigm, and performed grammaticality judgments. L1 interference was investigated by using syntactic structures that differ across Arabic and Hebrew.

To summarize, the results of the current investigation have the potential of expanding our understanding of cross-language interference in different-script bilinguals, the extent to which such interference management is similar across language domains, and the degree to which it might be modulated by L2 proficiency and domain general cognitive control.

2. Method

2.1. Participants

A total of 166 students at the University of Haifa participated in the study. Sixty-eight participants (20 males) were native Arabic speakers, who spoke Hebrew as an L2, and completed all experimental tasks. The native Arabic speaking participants had studied Hebrew in a formal setting beginning in the third grade, and at the time of testing were partially immersed in Hebrew. Their university courses were conducted in Hebrew, but most of them resided in Arabic speaking communities. Native Arabic speakers had also studied English beginning in the fourth grade.

In addition, 98 native Hebrew speakers (20 males) served as controls. Of these, 38 completed both experimental language tasks (Cross-Modal priming and Self-paced reading), 30 completed only the cross-modal task and 30 completed only the self-paced reading task. All native Hebrew speakers also completed the language history questionnaire and the non-verbal intelligence measure. Native Hebrew speakers had no knowledge of Arabic, but had studied English in a formal setting since the third grade. A full description of participants' language background is presented in Table 1. Participants had no history of learning disabilities or language impairment and were recruited through advertisements offering payment (approximately 10 USD per hour) or course-credit for participating. Participants signed an informed consent in which they approved their participation in the study.

2.2. Tasks and materials

2.2.1. Language proficiency assessment

2.2.1.1. Language history questionnaire. A Hebrew translation of the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007) was administered, evaluating the exposure, use and proficiency of all known languages.

2.2.1.2. Peabody picture vocabulary test. The Hebrew adaptation of the PPVT test (Solberg & Nevo, 1979, based on Dunn, 1965) was used to evaluate participants' Hebrew vocabulary knowledge. In this test the examiner said a word orally and participants chose one picture of four that matches the word.

2.2.1.3. University entrance Hebrew language test (YAEL). Native Arabic participants gave their consent to retrieve their scores on the Hebrew proficiency university entrance examination. This test is mandatory for non-native Hebrew speakers enrolling in higher education in Israel (parallel to the English TOEFL), and includes two parts: The first part includes multiple choice sub-tests of sentence completion, restatements and reading comprehension. The second part consists of a composition. Final scores range from 50 to 150. At the University of Haifa, incoming students must achieve a score of 120 or up to be admitted. Students receiving scores in the range of 90–119 are accepted on probation, and must re-sit the exam after their first year of study, and reach a score of 120.

2.2.2. Domain general executive function tests

2.2.2.1. Mental shifting ability. Examined by the task switching paradigm adopted from Prior and MacWhinney (2010). In this task participants are presented with red or green triangles and circles, and need to make color and shape judgments. The cue for the color task was a color gradient, and the cue for the shape task was a row of small black shapes. One response for each task (red and green for color, and circle or triangle for shape) was mapped to the right hand and the other to the left hand. The task includes single task and mixed task blocks, which included unpredictable switches between tasks. In the mixed blocks 50% of the trials were switch trials and 50% were repeat trials.

The task yields two measures of shifting abilities: Switching costs are the differences in performance between switch and non-switch trials in the mixed blocks; Mixing costs are the differences in performance between single-task trials in the single-task blocks and non-switch trials in the mixed blocks. Two joint measures, each incorporating costs in both RTs and accuracy were calculated – one for switching and one for mixing – using the bin scoring method, as elaborated in the results section (Hughes, Linck, Bowles, Koeth, & Bunting, 2014).

Table 1
Participant characteristics by language group.

	Native Arabic (SE)	Native Hebrew (SE)
Age*	21.85 (0.26)	25.97 (0.36)
Arabic self-rated proficiency	9.71 (0.08)	–
Arabic self rated use	56.84 (1.97)	–
Hebrew self-rated proficiency*	8.23 (0.15)	9.83 (0.03)
Hebrew self rated use*	35.65 (1.97)	83.35 (1.72)

Note: Self-rated proficiency scores were given on a scale from 1 to 10. Proficiency was averaged across oral and written comprehension and production. Self-Rated use was reported in percentage of the time participants used each language currently. Percentages to not reach 100%, because participants also reported using English part of the time.
* $p < 0.001$.

2.2.2.2. Inhibitory function. Inhibitory function was measured using two nonlinguistic tasks.

2.2.2.2.1. Numeric Stroop. Numeric Stroop was adopted from Hernandez et al. (2010). In each trial, participants were instructed to indicate by button press, as quickly and accurately as possible, how many items appear on the screen (range from 1 to 3). There were three experimental conditions: neutral (xxx), congruent (333), and incongruent (111), with trials presented in a random order. The interference effect was calculated as differences in performance between congruent and incongruent trials, using the bin scoring method (Hughes et al., 2014).

2.2.2.2.2. Simon arrows. Simon Arrows (also called spatial Stroop) adopted from Bialystok et al. (2008; used also by Mor et al., 2014). In each trial, an arrow pointing either left or right appeared on the screen. In the basic condition, the arrow appeared in the center of the screen, and participants were instructed to respond to the direction of the arrow as quickly and accurately as possible. In the conflict condition, the target arrows were presented on the left or right sides of the display, creating congruent trials when the direction and position corresponded, and incongruent trials when they are in conflict. Participants were instructed to press the button indicating the direction of the arrow irrespective of the position. The Simon effect was calculated as differences in performance between congruent and incongruent trials in the conflict block, using the bin scoring method (Hughes et al., 2014).

Finally, following the procedures of Bialystok et al. (2008; see also Mor et al., 2014), the reverse condition was a measure of response inhibition: the display was identical to the display in the basic condition with an arrow appearing in the center of the screen. However, the instructions were reversed – to press the response button in the direction opposite to the one indicated by the arrow. The Reverse effect was calculated as differences in performance between the basic and reverse blocks, using the bin scoring method (Hughes et al., 2014).

2.2.2.3. Memory tasks. In order to give a full characterization of executive functions, native Arabic speaking participants also completed a working memory task, namely the Operation span task adapted from Turner and Engle (1989). However, participants had very low accuracy rates on the mathematical operations (M = 62% correct), deeming the word span memory measure uninterpretable. Thus, working memory (updating) abilities were not included in the predictive models reported in the results section.

Finally, native Arabic participants completed two additional tasks: Raven's Progressive Matrices – a measure of nonverbal intelligence (Raven, Raven, & Court, 1976), and a phonological short-term memory measure, examined using a non-word repetition task (Shatil, 2002). Performance on both tasks did not correlate with performance on the experimental tasks or on executive function measures, and they are therefore not discussed further.

2.2.3. Linguistic experimental tasks

Two tasks were designed to probe cross-language activation and interference from L1 (Arabic) when performing a task in the L2 (Hebrew), for native Arabic speakers. For each of these tasks a group of native Hebrew speakers served as controls (see participant section above).

2.2.3.1. Cross modal priming. Participants listened to a Hebrew prime-word (recorded by a female native Hebrew speaker) through headphones, and were then presented with a written Hebrew target word on a computer screen. They were asked to decide by button press if the target word that they saw was semantically related to the auditory prime-word, by pressing the left button for 'no', and the right button for 'yes'.

On critical trials, the prime was a false-cognate in Hebrew and Arabic – it was phonologically similar in Arabic and Hebrew but did not have the same meaning. For example, "bread" in Hebrew is pronounced "*leh'em*", whereas in Arabic the word "*lah'em*" sounds very similar, but means "meat". Target words (e.g., Butcher shop, "*Itliz*" in Hebrew) were semantically related to the Arabic meaning of the false-cognate (meat) but were not related to the Hebrew meaning (bread), so the correct response in these trials was always "no". For each critical prime a matched control item was selected. Control primes did not share either meaning or phonology in Hebrew and Arabic (for example, "clothing" – '*beged*' in Hebrew and *'lebas'* in Arabic), and were also not semantically related to the target word (butcher shop), so the correct response is also "no" (for a full example, see Table 2).

Critical and control primes did not differ significantly in either length in phonemes ($t(41) = 0.18$, $p = 0.855$) ($M = 3.6$, $SD = 0.96$; $M = 3.6$, $SD = 0.91$; for critical and control primes, respectively) or lexical frequency in Hebrew (based on HebWaC corpus via Sketch-Engine, see Kilgarriff, Reddy, Pomikálek, & Avinesh, 2010; Kilgarriff et al., 2014) ($t(40) = 1.34$, $p = 0.187$) ($M = 54.2$, $SD = 65.9$; $M = 41.5$, $SD = 65.3$, for critical and control primes, respectively). In addition, semantic similarity judgments for the prime-target pairs (critical and control) were gathered from 10 native Hebrew speaking university students (who do not know Arabic) on a scale of one to seven, for course credit. To this end, two versions of a computerized questionnaire were created, such that each target word appeared once in each version, but across participants each target word was presented with both critical and control primes. There were no significant differences in the rated semantic similarity of the critical primes and the control primes with the target words ($t(41) = 1.39$, $p = 0.170$) ($M = 1.5$, $SD = 0.6$, $M = 1.4$, $SD = 0.6$ for critical and control primes, respectively).

In addition to critical items, two types of fillers were added to the list. First, an equal number of Arabic-Hebrew cognates were added to the list, such that phonological overlap was not an immediate cue for meaning competition. For example, the word "*yad*" that sounds similar in Arabic and in Hebrew, carries the meaning "hand" in both languages. The target words following cognate primes (e.g., knee) were semantically related to the auditory prime-word and thus, the correct answer in these trials was always 'yes'. For each cognate prime, a control prime that was equally semantically related to the target was selected.

Second, 78 filler pairs, in which primes were neither cognates nor false-cognates between Hebrew and Arabic, were added to the experiment in order to conceal the form overlap manipulation. In half of these filler pairs, the prime was semantically related to the target (e.g. chair-table) and in the other half it was semantically unrelated (e.g., lamp-socks).

Table 2
Example of false-cognate critical and control stimuli.

		Critical	Control
Prime	Pronunciation	*Lechem*	*Beged*
	Meaning in Hebrew	Bread	Clothing
	Meaning in Arabic	Meat	–
Target	Pronunciation	*Itliz*	
	Meaning in Hebrew	Butcher shop	
	Meaning in Arabic	–	

Two experimental lists were constructed, each completed by half of the participants in each group in the main study. Each version contained 162 stimuli, 42 experimental false-cognate items (21 critical primes and 21 control primes), 42 filler cognate items (21 critical primes and 21 control primes), and 78 phonologically unrelated fillers (39 semantically related and 39 semantically unrelated). Each participant saw each target word only once, but each target appeared with control and critical primes for different participants. Stimuli were presented in random order to each participant.

The experimental task began with instructions presented on the computer screen, followed by eight practice trials. In each trial, participants first heard the prime word through headphones, and 500 ms later the target was presented on the screen until a response was given. The next trial began after an ITI of 500 ms. The list was presented in two blocks, with a short break introduced between them.

2.2.3.2. L2 self paced reading and grammaticality judgment. Participants read sentences in Hebrew, using a self-paced reading paradigm, and performed a grammaticality judgment at the end of each sentence. Half of the sentences were grammatical in Hebrew, and half were not. Five grammatical structures that are shared between Hebrew and Arabic (i.e., Similar Condition), and 5 grammatical structures where the languages differ (i.e., Different Condition) were used in the experiment (see Table 3).

For each structure, 12 sentences were constructed, with a grammatical variation and an ungrammatical variation for each sentence (for a full list of stimuli, see Appendix A). Although the position of the critical word creating the grammatical violation varied, there were always at least two words preceding it, and two content words following it. Target sentences in the different conditions were constructed such that relying on Arabic grammar would lead to an error in judging the Hebrew sentence (for example see Table 4).

From these materials, two experimental lists were constructed. One list included six grammatical and six ungrammatical sentences from each of the ten structures, and the second list was comprised of the complementary sentences. Each list was completed by half of the participants in each group. Therefore, across participants, both variations of each sentence were presented an equal number of times, but no participant saw both the grammatical and the ungrammatical variation of the same sentence.

Prior to the beginning of the experimental task, participants were presented with written instructions in Hebrew, followed by a practice block including four sentences. Participants were given feedback on their performance in the practice block, but not during the experiment itself.

Throughout the task, sentences were presented to participants using a self-paced reading paradigm using E-prime software (Psychology Software Tools, Pittsburgh, PA). Each trial started with a fixation, and participants pressed the response box to initiate the sentence and progress through it word-by-word. All words

Table 3
Similar and different syntactic structures in Hebrew and Arabic.

Similar structures	1. Subject-verb person agreement in active forms 2. Singular-plural subject verb agreement 3. Verb tense marking 4. Noun-adjective agreement for nouns that share gender across Hebrew and Arabic 5. Adjective agreement with animate nouns
Different structures	6. Verb object agreement 7. Direct object article before concrete and abstract nouns 8. Subject-verb agreement in first person 9. Noun-adjective agreement for nouns that differ in gender across languages 10. Adjective agreement with inanimate nouns

appeared at the center of the screen, and the sentence-final word appeared with a period or a question mark. Following each sentence, the question "Is this sentence grammatical?" in Hebrew appeared in the middle of the screen, and participants responded by button press, as quickly and as accurately as possible. Participants had the opportunity to take a short break after every 40 sentences.

Sensitivity to violations was evaluated by comparing reading times on the word at which the violation became evident in the ungrammatical conditions (critical word 'n') to reading times on the analogous word in the grammatical conditions. The word following the critical word ('n + 1') was analyzed to capture spillover effects. Notably, the 'n' and 'n + 1' words in the grammatical and ungrammatical variations of the sentences were necessarily different. The sentence-final word was analyzed to investigate sentence wrap-up effects. Sensitivity (d prime) of the grammaticality judgements was also analyzed.

2.3. Procedure

Participants provided informed consent for participation. Native Arabic speakers completed all tasks, for a total of approximately two and half hours. Participants were given the option to take breaks as needed. Participants who did not manage to complete all the tasks in one session returned for a second session. The two experimental interference tasks, namely cross-modal priming and the self-paced reading tasks, were always separated by at least 2 non-linguistic tasks. As noted above, native Hebrew controls completed one or both of the experimental language tasks, the non-verbal IQ task, and filled out the LEAP-Q.

3. Results

3.1. Lexical interference: cross-modal priming

In order to examine cross language influences we conducted separate two-way repeated measures ANOVA on RTs and accuracy rates. Interference was probed in a two-way analysis with condition (False cognate, control) as a within participant variable and language group (Hebrew, Arabic) as a between participant factor. A parallel item analysis was conducted. Two critical false-cognate items were removed from the final analysis, because accuracy rates for these items were below 60% in the native Hebrew control group suggesting that these were perceived as semantically related.In the RT analysis, there was a significant main effect of condition [$F_1(1,126) = 10.538$, $p < 0.001$, $\eta_p^2 = 0.077$; $F_2(1,39) = 6.332$, $p < 0.05$, $\eta_p^2 = 0.140$]. Overall, RTs to targets following false-cognate primes were slower than to targets following control primes. In addition, there was a significant main effect of language group [$F_1(1,126) = 76.817$, $p < 0.001$, $\eta_p^2 = 0.379$; $F_2(1,39) = 417.42$, $p < 0.001$, $\eta_p^2 = 0.915$], as native Hebrew speakers responded faster overall than native Arabic speakers. Importantly, these main effects were qualified by a significant two-way interaction in both the subject and the item analyses [$F_1(1,126) = 4.515$, $p < 0.05$, $\eta_p^2 = 0.035$; $F_2(1,39) = 5.234$, $p < 0.05$, $\eta_p^2 = 0.118$]. Planned comparisons showed that this interaction was driven by the fact that native Arabic speakers responded more slowly to targets following a false-cognate prime in comparison to targets following a control prime ($t_1(59) = 2.697$, $p < 0.01$; $t_2(39) = 2.637$, $p < 0.05$), whereas the native Hebrew speakers responded equally fast to both types of targets ($t_1(67) = 1.682$, NS; $t_2(39) = 0.995$, NS) (see Fig. 1, panel A).The accuracy analysis also showed a significant main effect of Condition [$F_1(1,126) = 77.945$, $p < 0.001$, $\eta_p^2 = 0.382$; $F_2(1,39) = 10.252$, $p < 0.01$, $\eta_p^2 = 0.208$], because accuracy rates following false-cognate primes were lower than following control prime.

Table 4
Grammatical and ungrammatical sentences with similar and different structures across Hebrew and Arabic. The word creating the violation is starred and underlined.

	Grammatical variation (Hebrew)	**Ungrammatical variation (Hebrew)**
Similar Structure	They are very reliable (plural) people and can be trusted הם אנשים מאוד **אמינים** ואפשר לסמוך עליהם. **Arabic translation:** هم أشخاص صادقون جدًا ويمكن *الاعتماد* عليهم.	They are very reliable (singular) people and can be trusted הם אנשים מאוד **<u>אמין</u>*** ואפשר לסמוך עליהם. **Arabic translation:** هم أشخاص **<u>صادق</u>*** جدًا ويمكن *الاعتماد* عليهم.
Different Structure	I have pretty (plural) pictures from many periods and places יש לי תמונות **יפות** מכל מיני תקופות ומקומות. **Arabic translation:** لدي العديد من الصور **<u>الجميلات</u>*** من فترات وأماكن مختلفة.	I have pretty (singular) pictures from many periods and places יש לי תמונות **<u>יפה</u>*** מכל מיני תקופות ומקומות. **Arabic translation:** لدي العديد من الصور **الجميلة** من فترات وأماكن مختلفة.

There was again a significant main effect of language group [$F_1(1,126) = 152.877$, $p < 0.001$, $\eta_p^2 = 0.548$; $F_2(1,39) = 105.935$, $p < 0.001$, $\eta_p^2 = 0.731$], because overall accuracy rates of the native Hebrew speakers were higher than those of the native Arabic speakers. Importantly, as in the RT analysis, the two-way interaction was significant [$F_1(1,126) = 18.853$, $p < 0.001$, $\eta_p^2 = 0.13$; $F_2(1,39) = 6.505$, $p < 0.05$, $\eta_p^2 = 0.143$]. As can be seen in Fig. 1 (panel B), the difference between false cognate and control primes was almost 3 times larger for the native Arabic speakers (13%) than for the native Hebrew speakers (5%), even though planned comparisons showed that the effect was significant for both groups ($t_1(59) = 7.403$, $p < 0.001$; $t_2(39) = 3.103$, $p < 0.05$ for Arabic speakers; $t_1(67) = 4.344$, $p < 0.001$; $t_2(39) = 2.581$, $p < 0.05$ for Hebrew speakers).

The unexpected difference in accuracy between critical and control primes found for the native Hebrew speakers makes it more difficult to ascribe the difference found for native Arabic speakers to cross-language interference. The difference for native Hebrew speakers may reflect problems in the initial classification of pairs as related or unrelated. This initial classification was based on a 1–7 off-line rating scale, which may not account fully for the performance on the online experimental task. In order to further investigate and clarify this issue, we collected additional data from a new group of 20 native Hebrew speakers. Based on the performance of this group, we selected a subset of better matched items. Specifically, items where accuracy for both the critical and control primes fell below 90% were eliminated, leading to a final set of n = 30. We chose this cutoff because performance accuracy was very high (M = 98%, SD = 4.1%) – thus, we eliminated items that deviated from the mean by more than two standard deviations. Analysis with this better matched subset of items revealed that the difference in accuracy between critical and control items remained robust and highly significant for the native Arabic speakers ($t_1(59) = 4.7$, $p < 0.001$; $t_2(29) = 2.1$, $p < 0.05$), but was no longer significant for the original group of native Hebrew speakers ($t_1(66) = 1.1$, $p = 0.27$; $t_2 < 1$). The pattern for RTs remained the same as in the initial analysis.

3.2. Grammatical interference: self-paced reading

In order to examine L1 interference in L2 grammatical processing, we conducted three 3-way repeated measures ANOVAs on reading times (RTs) of the critical word (n), the n + 1 word and the sentence-final word. Interference was probed in an analysis with grammaticality (grammatical, ungrammatical) and similarity (similar structure, different structure) as within participant variables, and language group (Hebrew, Arabic) as a between participant variable. A parallel item analysis was conducted, using a 3-way repeated measures ANOVA. Interference was probed in an analysis with grammaticality (grammatical, ungrammatical) and L1 group (Hebrew, Arabic) as within item variables, and Similarity (similar structure, different structure) as a between item variable. Nine item pairs were omitted from analysis because of low accuracy (under 60%).

In addition, a 2-way repeated measures ANOVA was conducted on sensitivity scores (d′ prime) of the grammaticality judgments with similarity (similar, different) as a within participant variable, and language group (Hebrew, Arabic) as a between participant variable.

3.2.1. Critical word reading times

Two main effects were significant in both analyses: grammaticality [$F_1(1,129) = 234.377$, $p < 0.001$, $\eta_p^2 = 0.645$; $F_2(1,109) = 181.559$, $p < 0.001$, $\eta_p^2 = 0.625$], and language group [$F_1(1,129) = 110.563$, $p < 0.001$, $\eta_p^2 = 0.462$; $F_2(1,109) = 564.659$, $p < 0.001$, $\eta_p^2 = 0.838$]. The main effect of similarity was significant in the subject analysis [$F_1(1,129) = 19.603$, $p < 0.001$, $\eta_p^2 = 0.132$], but not the item analysis [$F_2(1,109) = 1.061$, *NS*]. Here, as in all remaining analyses, reading times of native Arabic speakers (reading in the L2) were longer than of native Hebrew speakers (reading in the L1). The effects of grammaticality and of similarity were modified by significant two-way interactions with language group.

The two-way interaction between grammaticality and language group was significant [$F_1(1,129) = 48.512$, $p < 0.001$, $\eta_p^2 = 0.273$, $F_2(1,109) = 46.429$, $p < 0.001$, $\eta_p^2 = 0.299$]. Reading times of critical words were longer in the ungrammatical than in grammatical sentences across both participant groups but this effect was larger in the native Arabic speakers ($t_1(68) = 13.254$, $p < 0.001$; $t_2(110) = 10.848$, $p < 0.001$) than in the native Hebrew speakers ($t_1(61) = 8.554$, $p < 0.001$; $t_2(110) = 12.651$, $p < 0.001$) (see Table 5).

The two-way interaction of similarity and language group was also significant [$F_1(1,129) = 26.804$, $p < 0.001$, $\eta_p^2 = 0.172$; $F_2(1,109) = 4.429$, $p < 0.05$, $\eta_p^2 = 0.039$]. This interaction was driven by the fact that for native Hebrew speakers critical-word reading times did not differ between similar and different structures ($t_1(61) = 1.045$, NS.; $t_2(110) = 0.949$, NS), whereas for native Arabic speakers critical-word reading times were shorter for different structures than for similar structures in the subject analysis ($t(68) = 5.389$, $p < 0.001$), but not in the item analysis ($t(110) = 1.505$, NS). All remaining interactions were not significant.

3.2.2. Spillover effects (word n + 1 reading times)

All three main effects were significant: grammaticality [$F_1(1,129) = 48.537$, $p < 0.001$, $\eta_p^2 = 0.273$; $F_2(1,109) = 40.305$, $p < 0.001$, $\eta_p^2 = 0.27$], similarity [$F_1(1,129) = 38.506$, $p < 0.001$, $\eta_p^2 = 0.230$; $F_2(1,109) = 7.355$, $p < 0.01$, $\eta_p^2 = 0.063$], and language group [$F_1(1,129) = 160.309$, $p < 0.001$, $\eta_p^2 = 0.554$; $F_2(1,109) = 418.935$, $p < 0.001$, $\eta_p^2 = 0.815$]. All three 2-way interactions were also significant [grammaticality by language group: $F_1(1,129) = 13.998$, $p < 0.001$, $\eta_p^2 = 0.098$; $F_2(1,109) = 13.737$, $p < 0.001$, $\eta_p^2 = 0.112$; similarity by language group: $F_1(1,129) = 31.511$, $p < 0.001$, $\eta_p^2 = 0.196$; $F_2(1,109) = 11.427$, $p < 0.001$, $\eta_p^2 = 0.095$; grammaticality by similarity: $F_1(1,129) = 20.231$, $p < 0.001$, $\eta_p^2 = 0.136$; $F_2(1,109) = 7.332$, $p < 0.001$, $\eta_p^2 = 0.063$]. Critically, these two-way interactions were further modified by a significant three-way interaction [$F_1(1,129) = 10.792$, $p < 0.001$, $\eta_p^2 = 0.077$; $F_2(1,109) = 6.17$, $p < 0.05$, $\eta_p^2 = 0.054$]. In order to specifically explore the effect of L1 interference, we examined the effect of grammatical overlap– comparing similar and different sentence structures – separately for the grammatical and ungrammatical conditions.

For grammatically correct sentences, there was a marginally significant interaction between language group and similarity, [$F_1(1,129) = 3.6$, $p = 0.06$, $\eta_p^2 = 0.03$; $F_2(1,109) = 2.782$, $p = 0.098$, $\eta_p^2 = 0.025$]. Planned comparisons revealed that native Hebrew speakers were equally fast in processing the n + 1 word in Hebrew structures regardless of whether they were shared with Arabic [$t < 1$, in both subject and item analysis]. In contrast, native Arabic speakers were marginally slower in reading the n + 1 word in Hebrew sentence structures that were not shared with Arabic in the subject analysis [$t(68) = 1.8$, $p = 0.07$], but this effect was not significant in the item analysis.

For ungrammatical sentences, a similar pattern of reading times emerged, but effects did reach statistical significance. There was a significant two-way interaction between language group and similarity, [$F_1(1,129) = 32.3$, $p < 0.001$, $\eta_p^2 = 0.2$; $F_2(1,109) = 349.172$,

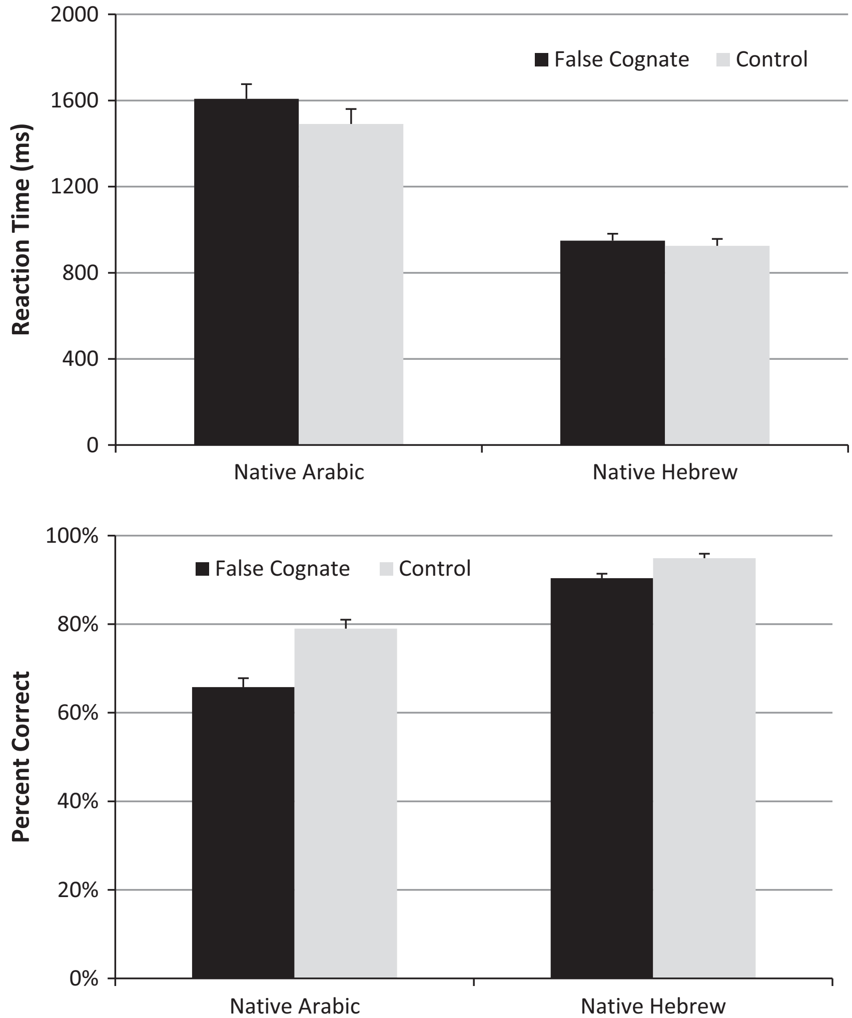

Fig. 1. RT (Panel A) and Accuracy (Panel B) in the false cognate and control conditions in the cross-modal semantic judgment task, by language group.

Table 5
Reading times in milliseconds (SEM) for the critical word, n + 1 word, and sentence final word, by grammaticality and structure similarity, by language group.

		Native Hebrew		Arabic-Hebrew bilinguals	
		Similar	Different	Similar	Different
Critical	Grammatical	536 (19)	523 (19)	1023 (41)	910 (35)
	Ungrammatical	643 (32)	673 (30)	1366 (57)	1254 (55)
n + 1	Grammatical	528 (19)	521 (19)	855 (34)	900 (32)
	Ungrammatical	564 (18)	584 (18)	935 (27)	1149 (39)
Final	Grammatical	673 (35)	762 (43)	1355 (90)	1416 (78)
	Ungrammatical	475 (20)	562 (26)	924 (67)	1132 (57)

$p < 0.001$, $\eta_p^2 = 0.762$]. Planned comparisons showed that for native Hebrew speakers there was a marginally significant difference in reading times for similar and different structures in the subject analysis only ($t_1(61) = 1.8$, $p = 0.08$; $t_2 < 1$]. However, native Arabic speakers were significantly slower in reading the n + 1 word in sentences with non-overlapping grammatical structures across the two languages in both analyses, [$t_1(68) = 6.9$, $p < 0.001$; $t_2(109) = 3.933$, $p < 0.001$].

As was the case in the lexical interference task, here again there is some evidence of the cross-language manipulation seeming to influence the performance of the native Hebrew speakers as well (even though of course they do not know Arabic). To address this issue, the new group of 20 native Hebrew speakers described above also completed the self-paced reading task. Based on their performance, we again selected a subset of better matched items. Performance accuracy of the native Hebrew speaking group was lower in this task (M = 91%, SD = 16%) than in the cross-modal task. In order to retain a sufficient number of items, we decided to employ a less stringent criterion than we used in the Cross-Modal priming task, and thus only eliminated sentences with accuracy rates of less than 80%. Then from the remaining stimuli we selected an equal number of sentences per condition, such that accuracy rates in the new group of participants were equivalent across the similarity conditions. As described below, analysis with this new matched set of 76 sentences resembled but was not identical to the pattern described above.

For grammatically correct sentences, there was now a significant interaction between language group and similarity, [$F_1(1,128) = 7.5$, $p = 0.007$, $\eta_p^2 = 0.06$; $F_2(1,74) = 5.46$, $p = 0.02$, $\eta_p^2 = 0.07$]. Planned comparisons revealed that native Hebrew speakers were equally fast in processing the n + 1 word in Hebrew structures regardless of whether they were shared with Arabic [$t < 1$, in both subject and item analysis]. In contrast, native Arabic speakers were significantly slower in reading the n + 1 word in Hebrew sentence structures that were not shared with Arabic in the subject analysis [$t_1(68) = 3.5$, $p = 0.001$] and marginally so in the item analysis [$t_2(74) = 1.8$, $p = 0.07$].

For the ungrammatical sentence we found a significant main effect of similarity ($F_1(1,128) = 52.8$, $p < 0.001$, $\eta_p^2 = 0.29$; $F_2(1,74) = 16.7$, $p < 0.001$, $\eta_p^2 = 0.19$) and an interaction with group ($F_1(1,128) = 49.5$, $p < 0.001$, $\eta_p^2 = 0.28$; $F_2(1,74) = 25.1$, $p < 0.001$, $\eta_p^2 = 0.25$). Planned comparisons show a significant effect of similarity for the native Arabic speakers ($t_1(68) = 7.9$, $p < 0.001$; $t_2(74) = 4.8$, $p < 0.001$) but this effect was now completely non-significant for the native Hebrew participants ($t_1 < 1$; $t_2 < 1$).[1] Thus, this improved analysis with better matched items shows significant disparities in processing similar and different grammatical structures for the native Arabic but not the native Hebrew speakers, increasing our confidence that cross-language interference is indeed the source of these effects.

3.2.3. Wrap up (sentence-final word reading times)

As in previous analyses, again all three main effects were significant: similarity, [$F_1(1,129) = 22.258$, $p < 0.001$, $\eta_p^2 = 0.147$; $F_2(1,109) = 4.602$, $p < 0.05$, $\eta_p^2 = 0.041$], language group, [$F_1(1,129) = 73.476$, $p < 0.001$, $\eta_p^2 = 0.363$; $F_2(1,109) = 299.964$, $p < 0.001$, $\eta_p^2 = 0.733$], and grammaticality, [$F_1(1,129) = 63.537$, $p < 0.001$, $\eta_p^2 = 0.33$; $F_2(1,109) = 50.334$, $p < 0.001$, $\eta_p^2 = 0.316$]. As has been reported in previous studies (Tokowicz & Warren, 2010), the effect of grammaticality was reversed in the final word, namely faster reading times in the final word of ungrammatical than grammatical sentences. The two-way interaction between grammaticality and language group was significant [$F_1(1,129) = 5.126$, $p < 0.05$, $\eta_p^2 = 0.038$; $F_2(1,109) = 5.112$, $p < 0.05$, $\eta_p^2 = 0.045$]. The effect of grammaticality was numerically larger for the native Arabic speakers than for the native Hebrew speakers, but was statistically significant in both groups ($ps < 0.001$) (see Table 5). All remaining interactions were not significant. The new analysis including the better matched subset of sentences yielded the same patterns of results, with the exception that in the item analysis the interaction between grammaticality and language group was only marginal, and not significant [$F_2(1,148) = 3.04$, $p = 0.083$, $\eta_p^2 = 0.02$].

3.2.4. D′ analysis on the grammaticality judgment accuracy

There was a significant main effect of syntactic similarity [$F(1,129) = 52.86$, $p < 0.001$, $\eta_p^2 = 0.484$]. Grammatical sensitivity in similar structures was higher than in different structure. The main effect of language group was also significant [$F(1,129) = 162.583$, $p < 0.001$, $\eta_p^2 = 0.558$]. As expected, native Hebrew speakers showed higher grammatical sensitivity than native Arabic speakers (see Fig. 2). Critically, there was a significant two-way interaction between similarity and language group [$F(1,129) = 22.307$, $p < 0.001$, $\eta_p^2 = 0.147$]. This interaction was driven by the fact that the effect of similarity, although statistically significant in both participant groups, was more than twice as large for native Arabic speakers ($t(68) = 11.109$, $p < 0.001$) than for native Hebrew speakers ($t(61) = 4.47$, $p < 0.001$) (see Fig. 2).

An analysis of d′ sensitivity performance based on the new, better matched set of materials yielded the exact same pattern of results, including a smaller but still significant effect of similarity for the native Hebrew speakers (interaction between similarity and language group [$F(1,128) = 4.18$, $p = 0.04$ $\eta_p^2 = 0.03$; effect for native Arabic speakers ($t(68) = 4.\ 64$, $p < 0.001$); effect for native Hebrew speakers ($t(60) = 3.0$, $p = 0.003$)].

3.3. Correlation and regression analyses

3.3.1. Correlations between lexical and grammatical interference management

The first issue we wished to examine was whether the ability to overcome L1 interference during L2 processing is general, i.e., shared across linguistic domains (lexicon and grammar). To this end, we calculated the correlations between measures of interference extracted from each of the tasks, across the group of Arabic-Hebrew bilinguals. From the cross-modal priming task, we focused on the interference effect in RT and accuracy. From the Self-Paced reading task, we selected the difference in reading times of the n + 1 word between the different vs. similar structures, collapsing across grammatical and ungrammatical sentences. We opted to collapse across both categories, because in both cases Arabic-Hebrew bilinguals showed reliably longer RTs for different structures than for similar structures, whereas native Hebrew speakers were not affected by similarity. In addition, we included the d′ prime measure from the different condition, as it indexes participants' ability to base their judgments solely on the L2 grammar, while overcoming interference from the L1 grammar. These two measures were selected because cross-language interference was most robust on these measures. Correlations across all these measures were found to be non-significant (all $r < 0.15$ all $p > 0.28$, see Fig. 3). Such a lack of correlation could stem from noisy measures, however, typical split-half correlations (even-odd) show reasonable reliability ($0.57 < r < 0.89$, all $ps > 0.0001$). Thus, we did not find support for the notion that the ability to overcome interference from the L1 when processing L2 is generalized across lexicon and grammar.

3.3.2. Predicting cross-language interference with L2 proficiency and executive function

We further examined whether the ability to overcome L1 interference might be related to general measures of L2 proficiency as well as domain-general executive function abilities, within the Arabic-Hebrew bilingual participants. Again, cross-language interference was indexed by the four dependent measures that yielded the most robust effects: Interference effect in RT and accuracy from the cross-modal priming task; the difference in n + 1 reading times between the different vs. similar structures from the self-paced reading task; the d′ prime measure from the different condition in the self-paced reading task.

Before conducting hierarchical regressions, we examined the correlations among the various L2 proficiency and executive function measures. Correlations among the L2 proficiency measures (Hebrew vocabulary, Hebrew university entrance exam, Hebrew proficiency self-rating were moderate (rs between 0.23 and 0.49, ps between 0.06 and 0.001, see Table 6).

For the executive function battery, all 10 standard EF effects were statistically significant (Simon congruency effect in RT and ACC, Simon reverse effect in RT and Accuracy, Stroop congruency effect in RT and accuracy, Switching costs and Mixing costs). Per-

[1] Analyzing reading times of the critical words in the subset of better-matched sentences described here yielded the exact same pattern of results as reported above in Section 3.2.1.

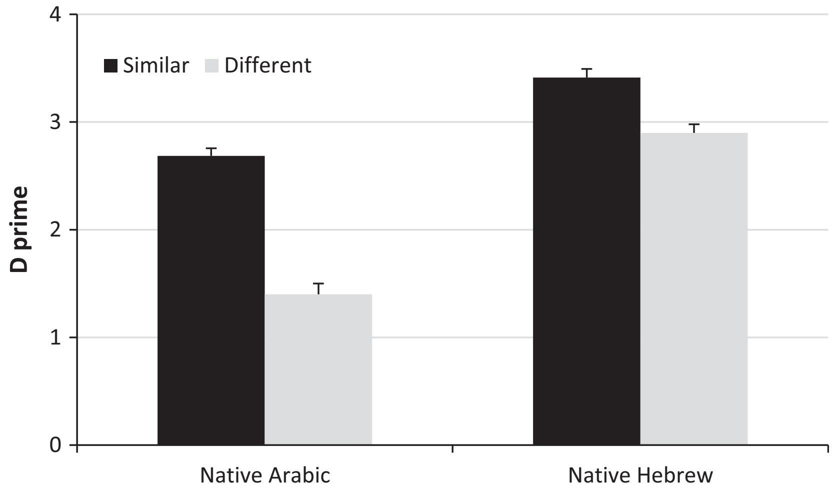

Fig. 2. D′ sensitivity measure for grammaticality judgments for similar and different structures, by language group.

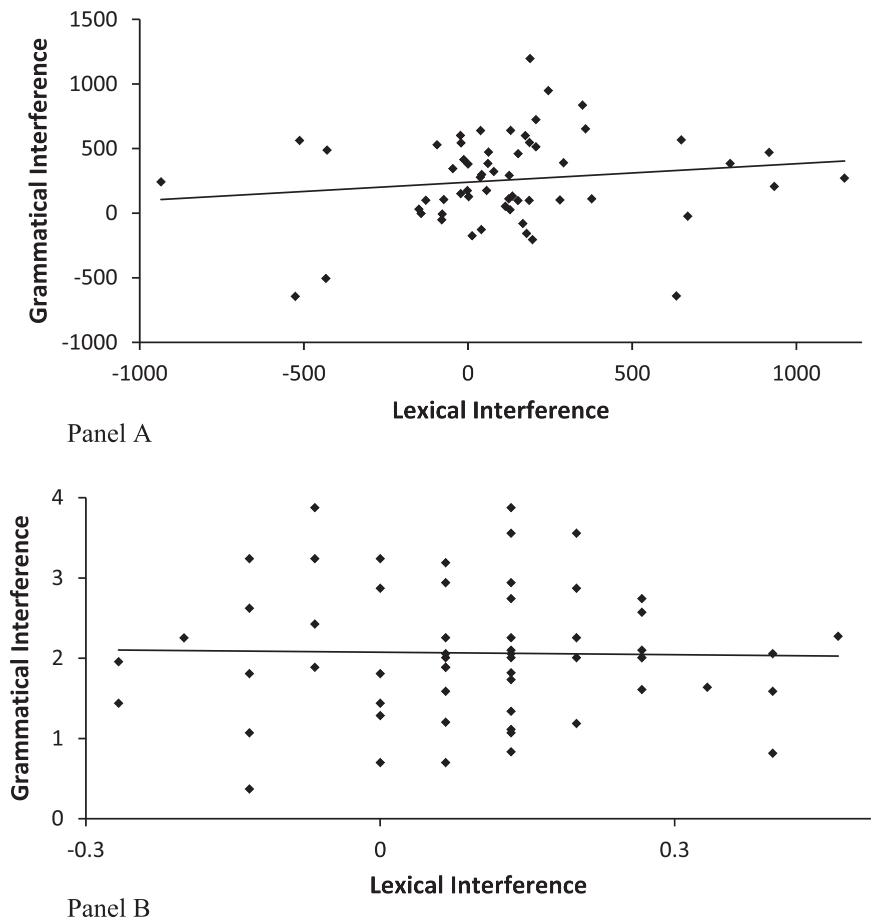

Fig. 3. Scatterplots of correlations between lexical and grammatical interference. Panel A: RT (Lexical interference: Difference between false cognates and control primes; Grammatical interference: Difference between n + 1 reading times for similar and different structures); Panel B: Accuracy (Lexical interference: Difference between false cognates and control primes; Grammatical interference: D′ sensitivity in grammaticality judgements of different structures.

Table 6
correlations matrix for L2 (Hebrew) proficiency scores.

	2	3
1. Hebrew vocabulary	0.492**	0.355**
2. University entrance exam		0.234±
3. Proficiency self-rating		

** $p < 0.01$.
± $p = 0.06$.

formance on the Simon, Stroop and task switching paradigms is presented in Tables 7 and 8.

Based on literature urging caution in using difference scores in prediction models, due to their potentially reduced reliability (e.g. Cronbach & Furby, 1970; Edwards, 2001; Hughes et al., 2014; Zimmerman, 2005) we decided to calculate bin scores for each of the EF effects. Bin scores create a combined measure for costs in performance in both RT and accuracy when comparing the more

Table 7
RT and accuracy (SEM) in the Spatial Simon and Numeric Stroop tasks.

	Spatial Simon		Numeric Stroop	
	RT	ACC	RT	ACC
Neutral	404 (6.3)	98% (0.3)	588 (9.4)	97.6% (0.3)
Congruent	448 (6.7)	99% (0.1)	541 (8.4)	99.2% (0.2)
Incongruent	501 (8.6)	96.6% (0.4)	619 (10.6)	91.6% (0.8)
Reverse	499 (12.0)	97.5% (0.4)	–	–
Congruency effect	53	2.4%	78	7.6%
Reverse effect	94	1.4%	–	–

Table 8
RT and accuracy (SEM) in the task switching paradigm.

	RT	ACC
Single	495 (12.7)	99% (0.1)
Repeat	965 (28.7)	96.9% (05)
Switch	1084 (29.8)	95.6% (04)
Mixing cost	470 (21.8)	2.1% (05)
Switching cost	119 (11.4)	1.3% (03)

Table 9
Correlation matrix for EF bin scores.

	2	3	4	5
1. Switch cost bin score	−0.15	0.03	0.15	−0.12
2. Mixing cost bin score		0.16	0.01	0.33*
3. Simon reverse bin score			0.11	0.13
4. Simon conflict bin score				0.03
5. Stroop bin score				

* $p < 0.01$.

challenging to the easier condition in each of the tasks. For each task, the RT on each "difficult" trial (switch trials in the task switching paradigm, and incongruent trials in the Simon and Stroop tasks) is subtracted from the participant's average RT for the "easy" trials. These differences for the entire group of participants are then rank-ordered, and placed into bins (ranging in value from 1 to 10). Each participant then received a "bin score" by summing up the number of his or her trials placed in each bin. In addition, error trials receive a score of 20 and are added to each participant's bin score. Thus, the bin score is a single measure that incorporates both RT and accuracy.

Higher bin scores indicate that a participant's performance suffered to a greater extent by the added cognitive burden in the difficult condition. To illustrate, a high bin score of the task switching cost reflects the fact that the participant had a greater cost in RT in the switch vs. the repeat condition (relative to other participants), and/or also tended to make more errors in the switch vs. the repeat condition (again, relative to other participants). Bin scores were calculated following the template set out by Hughes et al. (2014). As demonstrated in that paper, bin scores had improved reliability and validity relative to raw difference scores.

The 3 EF tasks used in the current study yielded 5 independent bin scores: switch cost (alpha Cronbach = 0.674), mixing cost (alpha Cronbach = 0.872), Stroop effect (alpha Cronbach = 0.784), Simon reverse effect (alpha Cronbach = 0.905) and Simon conflict effect (alpha Cronbach = 0.757). As can be seen by the alpha Cronbach coefficients, the reliability of these bin scores was sufficiently high to allow further assessment of the correlations among these measures, and between them and cross-language interference management.

Critically, as shown in Table 9, EF bin scores from different tasks (Simon, Stroop, and Switching) were mostly non-correlated for the current sample. Thus, although these tasks conceptually measure aspects of the shared construct of executive function, the current study aligns with previous reports of very low and non-significant correlations among them (Paap & Greenberg, 2013). Therefore, we were unable to reduce the dimensionality of the data (e.g., Principal Component Analysis did not identify separable factors).

We then conducted hierarchical regression analyses on the four measures of cross-language interference – namely, RT and accuracy costs for false cognates in the Cross-Modal task, difference in n + 1 reading times between different and similar structures in the Self-Paced reading task, and d′ for the grammaticality judgement task for different structures. We entered the L2 proficiency measures on the first step and the 5 EF bin scores on the second step. In these analyses, L2 proficiency significantly predicted performance in the d′ for different sentences from the grammaticality judgment task ($\Delta R^2 = 0.4$, $df = 60$, $p < 0.001$), such that higher L2 proficiency was associated with reduced cross-language interference. To reach a better understanding whether this finding indeed reflects a link between L2 proficiency and the specific ability to overcome L1 interference, we conducted an additional analysis, in which we entered participants' d′ score for similar sentences (reflecting their basic ability to perform grammaticality judgements in the L2) in the regression, only then followed by the L2 proficiency measures. In this analysis, L2 proficiency was again found to be a significant predictor of cross-language grammatical interference management ($\Delta R^2 = 0.31$, $df = 3$, $p < 0.001$), even after controlling for baseline grammatical knowledge in the L2. L2 proficiency did not significantly explain variability in performance in the remaining three measures of cross-language interference ($ps > 0.11$).

The contribution of the executive function measures to the regression models did not reach significance (for lexical interference effects in RT $p = 0.24$ and in accuracy $p = 0.06$, for grammatical interference effects in reading times $p = 0.35$ and in d′ $p = 0.72$).[2]

4. General discussion

The current study examined cross-language interference in proficient different-script bilinguals. In both lexicon and grammar, the current findings provide strong evidence of L1 interference during L2 processing. Specifically, lexical processing in Hebrew, the L2 of proficient Arabic-Hebrew bilinguals, was influenced by activation

[2] We also conducted hierarchical regression using the forward entry method, in which only significant predictors are retained in the model. These yielded the same results overall, with one exception. Namely, in predicting the accuracy effect in the lexical cross-modal task, the Simon conflict bin score was identified as explaining 8% (p = 0.038) of the variance (although with a negative beta value, such that higher interference effects in the Simon task were linked to smaller interference effects in accuracy in the lexical task). In addition, the mixing cost bin score was also identified as explaining an additional 7% of the variance (p = 0.045) (in this case smaller bin scores were associated with less interference in accuracy in the lexical task). Because of the theoretically unexpected direction of prediction, and in light of the relatively large number of regression models computed, we consider these to be spurious associations, and thus refrain from interpreting them further.

of the non-target language, Arabic. In a Hebrew semantic similarity judgment task, the non-target, Arabic, meaning of false-cognates hindered the performance of Arabic-Hebrew bilinguals, when compared to the performance of a control native Hebrew group, in both accuracy and reaction times. In addition, during an L2 (Hebrew) sentence reading task, Arabic-Hebrew bilinguals showed evidence of activation of the irrelevant grammatical features of the L1 (Arabic). As evident in both reading time measures and d′ in grammaticality judgments, Arabic-Hebrew bilinguals encountered more difficulty when processing Hebrew sentences whose grammatical structures differed between Hebrew and Arabic. These interference effects were manifest in the spillover and sentence final regions, but not on the critical word itself. It is possible that the slower time course of reading in the L2 (Copp, Drighe, & Duyck, 2015) results in delayed sensitivity to such grammatical manipulations (for a similar pattern see Tokowicz & Warren, 2010). This question should be further addressed in future research.

Notably, the native Hebrew speakers were also slightly less accurate when judging the grammaticality of sentences that differ in structure across Hebrew and Arabic, than those that share structure across languages. This difference cannot be explained by cross language influences, because the native Hebrew speaking participants did not know Arabic. Instead, it is most likely rooted in the fact that by definition unique grammatical structures were used in each of the conditions, and perhaps there are some inherent differences in their processing difficulty in Hebrew, possibly stemming from difference in frequency of occurrence of the constructions in the language. Critically, the effect was twice as large for Arabic-Hebrew bilinguals, and in addition only Arabic-Hebrew bilinguals, but not native Hebrew speakers, showed online sensitivity to the structural manipulation. Thus, we sustain that the performance of Arabic-Hebrew bilinguals was influenced, even if not exclusively, by interference from their L1, Arabic.

The cross-language effects demonstrated in the current study offer strong support for the notion of language non-selective activation (Degani & Tokowicz, 2010; Dijkstra, 2005; Kroll et al., 2014) even in different-script bilinguals. For such populations, orthographic stimuli are unambiguous with regards to language membership, and thus provide limited bottom-up activation of the non-target language. Despite these constraints, Arabic-Hebrew bilinguals were influenced by Arabic grammatical structures when reading visually presented sentences that were unambiguously in Hebrew. Thus, although the orthography is a strong and reliable cue to language membership, different-script bilinguals were unable to limit activation to the target (L2) language. Notably, we found robust L1 interference in pure L2 tasks that did not require dual-language activation. In previous studies, cross-language interference was documented for different-script bilinguals, but typically using tasks that inherently involved both language systems, such as translation recognition (Sunderman & Priya, 2012; but see Thierry & Wu, 2007, for analogous findings in a paradigm not involving both language systems). Moreover, the tasks in the current study were designed such that non-target language activation hindered performance in the target language. Nonetheless, it seems that L1 activation is inevitable even under such unfavorable conditions.

An additional goal of the present study was to probe to what degree cross-language interference management might be generalized across linguistic domains. To this end, the same bilingual participants performed both a lexical (semantic judgment) and a grammatical (sentence reading) task, which allowed us to examine possible correlations between participants' susceptibility to L1 interference when performing a pure L2 task in the lexical and the grammatical domains. Results showed no correspondence between interference management in the lexical and grammatical domain. This pattern suggests that cross-language interference management capabilities might develop independently for the lexico-semantic and grammatical domains, possibly influenced by the characteristics of each domain. Specifically, in the lexical domain managing cross-language interference might rely on mechanisms that are already in place to manage possible interference from within-language competitors (Marian & Spivey, 2003b). In contrast, such mechanisms might be less accessible in the grammatical domain, where fewer within-language interpretations normally exist for a given sentence.

However, the lexical and grammatical tasks used in the current study differed in many respects. Most importantly, the lexical task involved auditory stimuli, whereas the grammatical task was purely visual. In addition, the grammatical task more closely mimicked natural language processing, by presenting sentences rather than words in isolation. Thus, the fact that in the current study we did not find a significant correlation in interference management must be interpreted with some caution, as the measures of performance were quite different. It is possible that future studies using lexical and grammatical tasks designed to be more similar to each other might reveal some overlap in susceptibility to interference across the two domains.

Interestingly, L2 proficiency predicted cross-language interference management only in the grammatical domain, and only in the offline d′ measure. If indeed, as speculated above, interference management in the lexical domain relies on mechanisms built up through L1 experience, whereas interference management in the grammatical domain is trained more specifically through bilingual experience, it stands to reason that the latter but not the former would be related to overall L2 proficiency. However, the above described differences between the tasks might also contribute in some degree to these observed differences.

Finally, in the current study, domain general cognitive control did not significantly predict management of cross-language interference. This finding is especially important, because we probed the possible relation of these two cognitive constructs using multiple measures of control, and across two possible domains of cross-language interference (lexical and grammatical). Furthermore, the current study tested a relatively large number of participants (60), which arguably might more easily allow for detecting such a relationship.

Several recent studies did report significant statistical relations between domain general interference management and lexical cross-language interference management in bilinguals (Blumenfeld & Marian, 2013; Giezen et al., 2015; Mercier et al., 2014). However, the reported patterns of association are not consistent across studies. For example, whereas Blumenfeld and Marian (2013) link increased cognitive control to increased activation of cross-language competitors (though more efficient resolution of competition) in a visual world paradigm, Mercier et al. (2014) report an association in the opposite direction – namely, reduced activation of cross-language competitors linked to increased cognitive control (see also Giezen et al., 2015). In the current study, we failed to find a significant association between cognitive control and lexical cross-language interference, despite using several measures of cognitive control and a fairly large sample of participants. Although non-significant results may arise for different reasons, and their interpretation is not straight-forward, at the very least the current study suggests that future research should continue to investigate this issue. An accumulation of more empirical data will allow us to identify recurring patterns, and reach a better understanding of the possible theoretical and methodological conditions that lead to the mixed results across the extant studies.

Further, in the current study we also did not find a significant relation between domain general cognitive control and cross-language interference in the grammatical domain. This issue has received less attention in the literature, and to our knowledge the current study is the first to probe the possibility of such an association. Thus, the current findings do not support a link between cognitive control and cross-language interference across two language domains. We wish to put forth two possible explanations for this pattern of findings.

First, the current study investigated different-script bilinguals, whereas previous work focused on same-script bilinguals (Blumenfeld & Marian, 2013; Mercier et al., 2014; but see Giezen et al., 2015, for bi-modal bilinguals). As hypothesized in the introduction, language control and interference management might to some degree rely on different mechanisms in these different bilingual populations. To reiterate, cross-language activation might theoretically be reduce in different-script bilinguals, because the orthography provides a distinct cue to language membership (Gollan et al., 1997, but see Degani, Prior, & Tokowicz, 2011; Thierry & Wu, 2007). In addition, the two languages of different-script bilinguals may have developed differently than those of same-script bilinguals, with less overlap in activation between the two languages (Sunderman & Priya, 2012). Thus, although we found cross-language interference in different-script bilinguals, it still might be the case that due to the strong language-identity cue given in the orthographic representations, different-script bilinguals might rely to a lesser degree on domain general mechanisms for resolving cross-language interference.

A second, and by no means mutually exclusive, possible explanation stems from a consideration of the lack of convergence across the measures used in the current and previous research. As our own findings demonstrate, there was very little shared variance across different measures of cross-language interference on the one hand, and very little shared variance across different measures of cognitive control, on the other hand (Paap & Greenberg, 2013). Such a lack of convergence could inherently lead to variability in findings across studies – depending on the specific measures used for both theoretical constructs (cross-language interference, and cognitive control). Further, it might lead us to question the construct validity of cognitive control, and whether the various tasks indeed measure a single underlying mechanism.

Currently, we are unable to weigh the relative contribution of these two possible explanations, but future research examining diverse bilingual populations and measures should advance our understanding of these issues.

To summarize, we observed robust cross-language interference, in both lexicon and grammar, for proficient adult different-script bilinguals. Notably, interference management in these two domains seems to operate relatively independently, possibly relying on distinct mechanisms. Higher L2 proficiency was linked to improved interference management in the grammatical domain in an offline measure of accuracy but not in online measures of reading. In the lexical domain, however, L2 proficiency was not related to improved interference management in either accuracy or reaction times. Finally, cross-language interference management in the current study was not related to domain-general cognitive control. These findings provide insight into the underlying mechanisms of cross-language interference management in bilinguals, and suggest that, to some degree at least, interference management might not be shared across language domains, and across language and domain-general cognitive control.

Acknowledgements

The authors wish to thank Prof. Raphiq Ibrahim for assistance in developing stimuli for the grammatical task, and Walaa Hajajra, Zehavit Mohager, and Amal Ruhana for research assistance. The authors also thank Jared Linck and an anonymous reviewer for constructive comments on a previous version of the manuscript. AP was funded by EU-FP7 grant IRG-249163, and TD was funded by EU-FP7 grant CIG-322016.

Appendix A. Sentences used in the grammatical self-paced reading task, critical words are bolded

Structural overlap	Grammatical variation	Ungrammatical variation
Different	התקווה **היא** הבנים שלנו.	התקווה **הם** הבנים שלנו.
Different	החיים **הם** אגוז קשה.	החיים **הוא** אגוז קשה.
Different	הגמל **הוא** ספינת המדבר.	הגמל **היא** ספינת המדבר.
Different	דוח ועדת החקירה **הוא** עדות לכוח הדמוקרטיה.	דוח ועדת החקירה **היא** עדות לכוח הדמוקרטיה.
Different	אחת הסיבות לתאונות הדרכים **היא** חוסר שינה ועייפות שחווים נהגים רבים.	אחת הסיבות לתאונות הדרכים **הוא** חוסר שינה ועייפות שחווים נהגים רבים.
Different	הטלוויזיה **היא** התחליף לקולנוע.	הטלוויזיה **הוא** התחליף לקולנוע.
Different	הבאת ספר לימוד לכל שיעור **היא** תנאי הכרחי להצלחה.	הבאת ספר לימוד לכל שיעור **הוא** תנאי הכרחי להצלחה.
Different	הדם הוא הנפש שלנו.	הדם **היא** הנפש שלנו.
Different	האמצעים האמנותיים **הם** התבלין שביצירה.	האמצעים האמנותיים **הוא** התבלין שביצירה.
Different	ראיון עבודה **הוא** הזדמנו חד פעמית לשכנע את המראיין בכישרונותיך.	ראיון עבודה **היא** הזדמנות חד פעמית לשכנע את המראיין בכישרונותיך.
Different	השיר **הוא** פרי ייסורים רבים.	השיר **היא** פרי ייסורים רבים.
Different	שקריך **הם** הסיבה להרחקתך מחברת הילדים.	שקריך **היא** הסיבה להרחקתך מחברת הילדים.
Different	קניתי בחנות **את הספר** החשוב ביותר.	קניתי בחנות **הספר** החשוב ביותר.
Different	בחברה שלנו יש **את הבעיות** הקשות ביותר.	בחברה שלנו יש **הבעיות** הקשות ביותר.
Different	ראינו אתמול **את הסרט** החדש של ריצ'רד גיר.	ראינו אתמול **הסרט** החדש של ריצ'רד גיר.
Different	הוחלט לקיים **את האסיפה** ביום חמישי הקרוב.	הוחלט לקיים **האסיפה** ביום חמישי הקרוב.
Different	בגלל שהגשתי **את העבודה** באיחור המרצה לא קיבל אותה.	בגלל שהגשתי **העבודה** באיחור המרצה לא קיבל אותה.
Different	לא ביטלו **את הטיול** אף על פי שירדו גשמים עזים.	לא ביטלו **הטיול** אף על פי שירדו גשמים עזים.
Different	שמתי **את הלחם** במקרר החדש.	שמתי **הלחם** במקרר החדש.
Different	תרחץ בבקשה **את הפירות** בסבון ובמים חמים.	תרחץ בבקשה **הפירות** בסבון ובים חמים.
Different	שמתם במכונית **את כל** התיקים והחבילות.	שמתם במכונית **כל** התיקים והחבילות.
Different	ידעתי **את כל** התשובות במבחן המתמטיקה.	ידעתי **כל** התשובות במבחן במתמטיקה.

Appendix A (*continued*)

Structural overlap	Grammatical variation	Ungrammatical variation
Different	המורה לימד **את הכיתה** נושא חדש.	המורה לימד **הכיתה** נושא חדש.
Different	מילאתי **את הבקבוק** ושפכתי את המים על הילדים.	מילאתי את הבקבוק ושפכתי **המים** על הילדים.
Different	את **אמרת** לי כי עדיף לצאת לטיול מקודם בבוקר.	את **אמרתי** לי כי עדיף לצאת לטיול מקודם בבוקר.
Different	אימא אמרה לי שהיום את **אכלת** בבית של סבתא.	אימא אמרה לי שהיום את **אכלתי** בבית של סבתא.
Different	את **נכנסת** לי אל הלב.	את **נכנסתי** לי אל הלב.
Different	איך את **נמנעת** מנשנושים מיותרים לאורך זמן ממושך?	איך את **נמנעתי** מנשנושים מיותרים לאורך זמן ממושך?
Different	איך את **בנית** את המתחם החדש?	איך את **בניתי** את המתחם החדש?
Different	האם את **פנית** למחנכת הכיתה?	האם את **פניתי** למחנכת הכיתה?
Different	כל הכבוד לך **שתמכת** בנו לאורך כל השנה.	כל הכבוד לך **שתמכתי** בנו לאורך כל השנה.
Different	האם את **למדת** לבד למבחן במדעים?	האם את **למדתי** לבד למבחן במדעים?
Different	המורה אמרה לי שאת **רצית** ללמוד רפואה.	המורה אמרה לי שאת **רציתי** ללמוד רפואה.
Different	איפה את **למדת** גרפיקה והיסטוריה?	איפה את **למדתי** גרפיקה והיסטוריה?
Different	מתי את **נסעת** לאוניברסיטה היום?	מתי את **נסעת** לאוניברסיטה היום?
Different	שמעתי שאת **שמרת** על המסורת.	שמעתי שאת **שמרתי** על המסורת.
Different	אפשר לפתור בעיות **מתמטיות** מסובכות אם מתרכזים היטב.	אפשר לפתור בעיות מתמטיות **מסובכת** אם מתרכזים היטב.
Different	לקחתי מן הספרייה **העירונית** ספרים צבעוניים עם תמונות של בעלי חיים.	לקחתי מן הספרייה העירונית ספרים **צבעוני** עם תמונות של בעלי חיים.
Different	אפשר למצוא הרבה אבנים **גדולות** בחצר הבית שלנו.	אפשר למצוא הרבה אבנים **גדולה** בחצר הבית שלנו.
Different	הודיעו שהרוחות **החזקות** צפויות להימשך גם היום.	הודיעו שהרוחות **החזקה** צפויות להימשך גם היום.
Different	רעידת האדמה גרמה נזקים **רבים** ואבדות בנפש.	רעידת האדמה גרמה נזקים **רב** ואבדות בנפש.
Different	לדעתי אפשר למצוא מתכונים למאכלים **טעימים** באתר הזה.	לדעתי אפשר למצוא מתכונים למאכלים **טעים** באתר הזה.
Different	יש המון תוכניות **מעניינות** על מדע והיסטוריה.	יש המון תוכניות **מעניינת** על מדע והיסטוריה.
Different	יש לי הרבה תמונות **יפות** מכל מיני תקופות ומקומות.	יש לי הרבה תמונות **יפה** מכל מיני תקופות ומקומות.
Different	אני אוהב לארגן מסיבות **יוקרתיות** לעיתים קרובות.	אני אוהב לארגן מסיבות **יוקרתית** לעיתים קרובות.
Different	זוג הנאהבים ספרו זה לזו על חוויותיהם **המרגשות** בשנים האחרונות.	זוג הנאהבים סיפרו זה לזו על חוויותיהם **המרגשת** בשנים האחרונות.
Different	העצמות שלי עדיין **חלשות** למרות שאני שותה חלב כל בוקר.	העצמות שלי עדיין **חלשה** למרות שאני שותה חלב כל בוקר.
Different	יש להקפיד על שימוש במשפטים **פשוטים** בעת כתיבת עבודות אקדמיות.	יש להקפיד על שימוש במשפטים **פשוט** בעת כתיבת עבודות אקדמיות.
Different	קניתי אתמול כובע **כחול** בשוק העירוני.	קניתי אתמול כובע **כחולה** בשוק העירוני.
Different	לא כדאי לשתול עץ **גדול** בגינה קטנה שלא יסתיר את הבית והחצר.	לא כדאי לשתול עץ **גדולה** בגינה קטנה שלא יסתיר את הבית והחצר.
Different	יש לי משחק קלפים **מדהים** שמיועד לשני שחקנים או יותר.	יש לי משחק קלפים **מדהימה** שמיועד לשני שחקנים או יותר.
Different	לפני כ- 10 חודשים קניתי רכב **חדש** בסוכנות המרכזית.	לפני כ- 10 חודשים קניתי רכב **חדשה** בסוכנות המרכזית.
Different	קניתי שולחן אוכל **גדול** שנפתח עד לאורך של ארבעה מטרים.	קניתי שולחן אוכל **גדולה** שנפתח עד לאורך של ארבעה מטרים.
Different	צריך לשים לב אם הדלת **נעולה** כשיוצאים מהבית.	צריך לשים לב אם הדלת **נעול** כשיוצאים מהבית.
Different	היא גרה בכפר **קטן** ליד נצרת.	היא גרה בכפר **קטנה** ליד נצרת.
Different	איזה רעיון **נפלא** קראתי אתמול בעיתון.	איזה רעיון **נפלאה** קראתי אתמול בעיתון.
Different	אני מאחלת לכולם המשך לילה **טוב** והשכמה ליום חדש מחר בבוקר.	אני מאחלת לכולם המשך לילה **טובה** והשכמה ליום חדש מחר בבוקר.
Different	למרות שהארון **מלא** בגדים, אני תמיד מרגישה שאין לי מה ללבוש.	למרות שהארון **מלאה** בגדים אני תמיד מרגישה שאין לי מה ללבוש.
Different	הבטן שלי **נפוחה** אחרי כל ארוחה.	הבטן שלי **נפוח** אחרי כל ארוחה.
Different	מומלץ לתלות וילון **צבעוני** בחדר הילדים.	מומלץ לתלות וילון **צבעונית** בחדר הילדים.
Similar	דחפור ענק **הרס** את הבית הישן.	דחפור ענק **הרסה** את הבית הישן.
Similar	התלמידה החרוצה **מכינה** את שיעורי הבית באופן יומיומי.	התלמידה החרוצה **מכין** את שיעורי הבית באופן יומיומי.
Similar	אחותי הגדולה **מלמדת** עברית בחטיבת הביניים.	אחותי הגדולה **מלמד** עברית בחטיבת הביניים.
Similar	הסוחר הגבוה **העמיס** את הסחורה שלו על גג הרכב.	הסוחר הגבוה **העמיסה** את הסחורה שלו על גג הרכב.
Similar	הנער הזה **מפטפט** הרבה באמצע השיעור.	הנער הזה **מפטפטת** הרבה באמצע השיעור.
Similar	יועצת בית הספר **דיווחה** על המקרה לחוקרת הנוער.	יועצת בית הספר **דיווח** על המקרה לחוקרת הנוער.
Similar	הרופאה שלי **ביקשה** ממני לעשות בדיקות דם.	הרופאה שלי **ביקש** ממני לעשות בדיקות דם.
Similar	הסטודנט העצבני **שילם** את החוב במזכירות.	הסטודנט העצבני **שילמה** את החוב במזכירות.
Similar	השופט העסוק **דחה** את הישיבה לשבוע הבא.	השופט העסוק **דחתה** את הישיבה לשבוע הבא.
Similar	הילד הקטן **אכל** את הכריך בתאבון רב.	הילד הקטן **אכלה** את הכריך בתאבון רב.
Similar	למזלי אמא שלי **מכירה** את כל החברים שלי.	למזלי אמא שלי **מכיר** את כל החברים שלי.
Similar	התינוק הרעב **מתעורר** לאכול וחוזר לישון.	התינוק הרעב **מתעוררת** לאכול וחוזר לישון.
Similar	המתאגרפים **קיבלו** עידוד מילדיהם ומהקהל.	המתאגרפים **קיבל** עידוד מילדיהם ומהקהל.
Similar	האורח האחרון **הגיע** בשעת לילה מאוחרת.	האורח האחרון **הגיעו** בשעת לילה מאוחרת.
Similar	אנשים רבים **הלכו** לים בשל החום הכבד אשר שרר בארץ.	אנשים רבים **הלך** לים בשל החום הכבד אשר שרר בארץ.
Similar	הנער הצעיר **יצא** מביתו בעכו לכיוון צפת.	הנער הצעיר **יצאו** מביתו מעכו לכיוון צפת.
Similar	חוקרי התעופה **בדקו** את הממצאים מאתר ההתרסקות.	חוקרי התעופה **בדק** את הממצאים מאתר ההתרסקות.
Similar	רופא הילדים לא **יבוא** היום למרפאה.	רופא הילדים לא **יבואו** היום למרפאה.
Similar	החקירה מצאה שהאחראים **תיפקדו** היטב בזמן המשבר.	החקירה מצאה שהאחראים תיפקד היטב בזמן המשבר.
Similar	הארנבים הפחדנים לא **העיזו** להתקרב אלינו.	הארנבים הפחדנים לא **העיז** להתקרב אלינו.
Similar	הרעש שעושים הילדים לא **ייפסק** עד שהם ילכו לישון.	הרעש שעושים הילדים לא **ייפסקו** עד שהם ילכו לישון.
Similar	דברי המועמד **מצאו** חן בעיניי מאד.	דברי המועמד **מצא** חן בעיניי מאד.
Similar	נכתב בעיתון שהם **מרגלים** לטובת האויב.	נכתב בעיתון שהם **מרגל** לטובת האויב.

(*continued on next page*)

Appendix A (*continued*)

Structural overlap	Grammatical variation	Ungrammatical variation
Similar	לפני שהנהג **יצא** לדרך שאלתי אותו מה השעה.	לפני שהנהג **יצאו** לדרך שאלתי אותו מה השעה.
Similar	אתמול אנחנו **התחלנו** ללמוד בשעה שמונה בבוקר.	אתמול אנחנו **מתחילים** ללמוד בשעה שמונה בבוקר.
Similar	רציתי לשאול אותך למה אתמול **רצת** בשעה חמש בערב?	רציתי לשאול אותך למה אתמול **תרוץ** בשעה חמש בערב?
Similar	לפני חודש חנה **השכירה** חדרים לסטודנטים.	לפני שבוע חנה **משכירה** חדרים לסטודנטים.
Similar	לפני חודשיים רמי **ביטל** את הפגישה עם מר לוי.	לפני חודשיים רמי **מבטל** את הפגישה עם מר לוי.
Similar	בשבוע שעבר שלומית ושלמה **התחתנו** באירוע גדול ומפואר.	בשבוע שעבר שלומית ושלמה **תתחתנו** באירוע גדול ומפואר.
Similar	מחר בערב אנחנו **נהיה** אצל חברים טובים שלנו.	מחר בערב אנחנו **היינו** אצל חברים טובים שלנו.
Similar	אתמול יעקב **סיפר** לנו על המשפחה שלו.	אתמול יעקב **יספר** לנו על המשפחה שלו.
Similar	רק אתמול רונן **הבין** שהוא לא ראה את ההורים שלו כבר שבוע.	רק אתמול רונן **יבין** שהוא לא ראה את ההורים שלו כבר שבוע.
Similar	בעוד יומיים שלושת המדענים **יקבלו** פרס נובל על התרומה שלהם.	בעוד יומיים שלושת המדענים **קיבלו** פרס נובל על התרומה שלהם.
Similar	עבר כבר שבוע מאז שהוא **שלח** את המכתב והוא עוד לא הגיע.	עבר כבר שבוע מאז שהוא **ישלח** את המכתב והוא עוד לא הגיע.
Similar	כבר בשנות השישים **הבחינו** מדענים בפגיעה בשכבת האוזון.	כבר בשנות השישים **יבחינו** מדענים בפגיעה בשכבת האוזון.
Similar	בשיעור הפסנתר הבא רותי **תנגן** את היצירה המסובכת.	בשיעור הפסנתר הבא רותי **ניגנה** את היצירה המסובכת.
Similar	לפי דעתי עיניים **ירוקות** הן היפות ביותר.	לפי דעתי עיניים **ירוקים** הן היפות ביותר.
Similar	הדרך להצלחה **ארוכה** ודורשת השקעה רבה.	הדרך להצלחה **ארוך** ודורשת השקעה רבה.
Similar	אני מקבל עצה **טובה** מפיו של אדם מנוסה.	אני מקבל עצה **טוב** מפיו של אדם מנוסה.
Similar	שמענו מוסיקה **רועשת** מתוך הכפר הקרוב.	שמענו מוסיקה **רועש** מתוך הכפר הקרוב.
Similar	בשל רוחות **חזקות** נסגר שדה התעופה.	בשל רוחות **חזקים** נסגר שדה התעופה.
Similar	ביליתי יומיים **נהדרים** בירושלים עם החברים שלי.	ביליתי יומיים **נהדרות** בירושלים עם החברים שלי.
Similar	שמירה על שיניים **לבנות** מבטיחה חיוך יפה.	שמירה על שיניים **לבנים** מבטיחה חיוך יפה.
Similar	אני מעדיפה להשתמש בעיפרון **חדש** לכתוב תרגילים במחברת.	אני מעדיפה להשתמש בעיפרון **חדשה** לכתוב תרגילים במחברת.
Similar	אני אוהבת יונים **לבנות** יותר מאשר שחורות.	אני אוהבת יונים **לבנים** יותר מאשר שחורות.
Similar	שמעתי שבשנתיים **האחרונות** נשים יוצאות יותר לעבודה.	שמעתי שבשנתיים **האחרונים** נשים יוצאות יותר לעבודה.
Similar	הנהיגה בחורף מאוד **מסוכנת** ולכן יש לנסוע לאט.	הנהיגה בחורף מאוד **מסוכן** ולכן יש לנסוע לאט.
Similar	בערים **מעורבות** מדברים בשפות שונות.	בערים **מעורבים** מדברים בשפות שונות.
Similar	דני פגש חברים **ותיקים** אצל יוסף בחצר.	דני פגש חברים **ותיק** אצל יוסף בחצר.
Similar	המורות האלה **ותיקות** בתחום הלשון וההבעה.	המורות האלה **ותיקה** בתחום הלשון וההבעה.
Similar	ריהוט שמיוצר על ידי נגרים **מקצועיים** הוא בדרך כלל יקר יותר.	ריהוט שמיוצר על ידי נגרים **מקצועי** הוא בדרך כלל יקר יותר.
Similar	חנה ודינה הן הסטודנטיות הכי **חרוצות** בפקולטה להנדסת חשמל.	חנה ודינה הן הסטודנטיות הכי **חרוצה** בפקולטה להנדסת חשמל.
Similar	האורחים **המקסימים** הגיעו בדיוק בזמן.	האורחים **המקסים** הגיעו בדיוק בזמן.
Similar	הם אנשים מאוד **אמינים** ואפשר לסמוך עליהם.	הם אנשים מאוד **אמין** ואפשר לסמוך עליהם.
Similar	בקופת-חולים עובדים רופאים **צעירים** שהגיעו מאיטליה.	בקופת-חולים עובדים רופאים **צעיר** שהגיעו מאיטליה.
Similar	במשרד הפנים העולים **החדשים** קיבלו תעודת זהות.	במשרד הפנים העולים **החדש** קיבלו תעודת זהות.
Similar	ילדים **מפונקים** חושבים שהכל צריך לבוא להם בקלות.	ילדים **מפונק** חושבים שהכל צריך לבוא להם בקלות.
Similar	כיום יש הוכחה לכך שנשים **אמינות** יותר מגברים.	כיום יש הוכחה לכך שנשים **אמינה** יותר מגברים.
Similar	חברים **טובים** הם כמו משפחה.	חברים **טוב** הם כמו משפחה.
Similar	כמה תלמידים **שובבים** מפריעים למורה לנהל את השיעור כראוי.	כמה תלמידים **שובב** מפריעים למורה לנהל את השיעור כראוי.

Appendix B. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cognition.2017.04.006.

References

Bialystok, E., Craik, F. I. M., & Luk, G. (2008). Cognitive control and lexical access in younger and older bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition, 34*, 859–873.

Blumenfeld, H. K., & Marian, V. (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals. *Journal of Cognitive Psychology, 25*, 547–567.

Chambers, C. G., & Cooke, H. (2009). Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning, Memory and Cognition, 35*, 1029–1040.

Copp, U., Drighe, D., & Duyck, W. (2015). Eye movement patterns in natural reading: A comparison of monolingual and bilingual reading of a novel. *PLoS ONE*. http://dx.doi.org/10.1371/journal.pone.0134008.

Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological Bulletin, 74*, 68–80.

Degani, T., Prior, A., & Tokowicz, N. (2011). Bidirectional transfer: The effect of sharing a translation. *Journal of Cognitive Psychology, 23*, 18–28.

Degani, T., & Tokowicz, N. (2010). Semantic ambiguity within and across languages: An integrative review. *Quarterly Journal of Experimental Psychology, 63*, 1266–1303.

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism. Psycholinguistic approaches* (pp. 179–201). Oxford: Oxford University Press.

Dijkstra, T., & Van Heuven, W. J. B. ((2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition, 5*, 175–197.

Dimitropoulou, M., Duñabeitia, J. A., & Carreiras, M. (2011). Phonology by itself: Masked phonological priming effects with and without orthographic overlap. *Journal of Cognitive Psychology, 23*, 185–203.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589–630). Oxford, UK: Blackwell Publishing Ltd..

Dunn, L. H. (1965). *Manual for peabody picture vocabulary test*. Minnesota: American Guidance Services Inc..

Dussias, P. E. (2003). Syntactic ambiguity resolution in L2 learners: Some effects of bilinguality on L1 and L2 processing strategies. *Studies in Second Language Acquisition, 25*, 529–557.

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods, 4*, 265–287.

Festman, J., Rodriguez-Fornells, A., & Münte, T. F. (2010). Individual differences in control of language interference in late bilinguals are mainly related to general executive abilities. *Behavioral and Brain Functions, 6*, 5.

Friesen, D. C., & Jared, D. (2012). Cross-language phonological activation of meaning: Evidence from category verification. *Bilingualism: Language and Cognition, 15*, 145–156.

Giezen, M. R., Blumenfeld, H. K., Shook, A., Marian, V., & Emmorey, K. (2015). Parallel language activation and inhibitory control in bimodal bilinguals. *Cognition, 141*, 9–25.

Gollan, T. H., Forster, K., & Frost, R. (1997). Translation priming with different scripts: Masked priming with cognates and noncognates in Hebrew-English bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1122–1139.

Guo, T., Liu, H., Misra, M., & Kroll, J. F. (2011). Local and global inhibition in bilingual word production: fMRI evidence from Chinese-English bilinguals. *Neuroimage, 56*, 2300–2309.

Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition, 1*, 213–229.

Hernandez, M., Costa, A., Fuentes, L. J., Vivas, A., & Sebastian-Galles, N. (2010). The impact of bilingualism on the executive control and orienting networks of attention. *Bilingualism: Language and Cognition, 13*, 315–325.

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition, 106*, 501–511.

Hughes, M. M., Linck, J. A., Bowles, A. R., Koeth, J. T., & Bunting, M. F. (2014). Alternatives to switch-cost scoring in the task switching paradigm: Their reliability and increased validity. *Behavior Research Methods, 46*, 702–721.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., ... Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography, 1*, 7–36. http://dx.doi.org/10.1007/s40607-014-0009-9.

Kilgarriff, A., Reddy, S., Pomikálek, J., & Avinesh, P. V. S. (2010). A corpus factory for many languages. In *LREC workshop on web services and processing pipelines, Malta, May 2010* (pp. 2010).

Kim, J., & Davis, C. (2003). Task effect in masked cross-script translation and priming. *Journal of Memory and Language, 49*, 484–499.

Kotz, S. A. (2010). A critical review of ERP and fMRI evidence on L2 syntactic processing. *Brain & Language, 109*, 68–74.

Kroll, J. F., Bobb, S. C., & Hoshino, N. (2014). Two languages in mind: Bilingualism as a tool to investigate languages, cognition, and the brain. *Current Directions in Psychological Science, 23*, 159–163.

Lagrou, E., Hartsuiker, R. J., & Duyck, W. (2013). The influence of sentence context and accented speech on lexical access in second-language auditory word recognition. *Bilingualism: Language & Cognition, 16*, 508–517.

Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 381–390.

Linck, J. A., Hoshino, N., & Kroll, J. F. K. (2008). Cross-language lexical processes and inhibitory control. *Mental Lexicon, 3*, 349–374.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language. Evidence for the role of inhibition in second-language learning. *Psychological Science, 20*, 1507–1515.

Linck, J. A., Schwieter, J. W., & Sanderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition, 15*, 651–662.

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). Oxford: Oxford University Press.

Marian, V., Blumenfeld, H., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*, 940–967.

Marian, V., & Spivey, M. (2003a). Bilingual and monolingual processing of competing lexical items. *Applied Psycholinguistics, 24*(2), 173–193.

Marian, V., & Spivey, M. (2003b). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition, 6*(2), 97–115.

Mercier, J., Pivneva, I., & Titone, D. (2014). Individual differences in inhibitory control relate to bilingual spoken word processing. *Bilingualism: Language and Cognition, 17*, 89–117.

Mor, B., Yitzhaki-Amsalem, S., & Prior, A. (2014). The joint effect of bilingualism and ADHD on executive functions. *Journal of Attention Disorders, 19*, 527–541.

Morford, J. P., Kroll, J. F., Piñar, P., & Wilkinson, E. (2014). Bilingual word recognition in deaf and hearing signers: Effects of proficiency and language dominance on cross-language activation. *Second Language Research, 30*, 251–271.

Nakayama, M., Verdonschot, R. G., Sears, C. R., & Lupker, S. J. (2014). The masked cognate translation priming effect for different-script bilinguals is modulated by the phonological similarity of cognate words: Further support for the phonological account. *Journal of Cognitive Psychology, 26*, 714–724.

Nitschke, S., Kidd, E. J., & Serratrice, L. (2010). First language transfer and long-term structural priming in comprehension. *Language and Cognitive Processes, 25*, 94–114.

Paap, K., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology, 66*, 232–258.

Pivneva, I., Mercier, J., & Titone, D. (2014). Executive control modulates cross-language lexical activation during L2 reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*, 787–796.

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish-English and Mandarin-English bilinguals. *Journal of International Neuropsychological Society, 17*, 682–691.

Prior, A., & MacWhinney, B. (2010). A bilingual advantage in task switching. *Bilingualism: Language and Cognition, 13*, 253–262.

Raven, J., Raven, J. C., & Court, J. H. (1976). *Raven's progressive matrices*. Oxford: Oxford Psychologists Press.

Roberts, L. (2012). Individual differences in second language sentence processing. *Language Learning, 62*, 172–188.

Roberts, L., Gullberg, M., & Indefrey, P. (2008). Online pronoun resolution in L2 discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition, 30*, 333–357.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research, 24*, 397–430.

Shatil, E. (2002). *The Shatil test for early detection of specific disabilities and of general language impairments in the acquisition of reading and spelling: The comprehensive evaluation*. Kiryat Bialik: Ach Inc..

Solberg, S., & Nevo, B. (1979). Preliminary steps towards an Israeli standardization of the Peabody Test. *Megamot, 3*, 407–413 (In Hebrew).

Sunderman, G. L., & Priya, K. (2012). Translation recognition in highly proficient Hindi-English bilinguals: The influence of different scripts but connectable phonologies. *Language and Cognitive Processes, 27*, 1265–1285.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences, 104*, 12530–12535.

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar. *Studies in Second Language Acquisition, 27*, 173–204.

Tokowicz, N., & Warren, T. (2010). Beginning adult L2 learners' sensitivity to morphosyntactic violations: A self-paced reading study. *European Journal of Cognitive Psychology, 22*, 1092–1106.

Tuninetti, A., Warren, T., & Tokowicz, N. (2014). Cue strength in second-language processing: An eye-tracking study. *The Quarterly Journal of Experimental Psychology, 68*, 568–584.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154.

Van Hell, J. G., & Tanner, D. (2012). Second language proficiency and cross-language lexical activation. *Language Learning, 62*, 148–171.

Zimmerman, D. W. (2005). A note on interpretation of formulas for the reliability of differences. *Journal of Educational Measurement, 31*, 143–147.