Appendix A. Individual measures of linguistic and cognitive ability

Participants were administered a number of measures to assess language proficiency in the L1 and L2, as well as tasks designed to ascertain individual differences in cognitive control. These are described below.

1. Measures of language proficiency and language ability
1.1. Language experience and proficiency questionnaire
To assess linguistic proficiency and background in the L1 and the L2, participants completed an abridged version of the LEAP-Q (Marian, Blumenfeld, & Kaushanskaya, 2007). Through the LEAP-Q, data about an individual's L2 learning background were collected (Onset of Acquisition, weekly exposure, immersion) and other measures relevant to assess learning ability (age, level of education).

1.2. L2 proficiency
General English proficiency was measured by administering an abridged version of the Michigan English Language Institute College English Test (MELICET). Participants also completed a vocabulary test to measure general lexical knowledge, and a multiple-choice test to assess knowledge of the individual words employed in the experiment. High scores in the vocabulary tests demonstrated sufficient knowledge of single words (Unrelated mean: 92%; SD: 4.9; L1-interference mean: 91%; SD: 6.3), with no significant differences between the two groups ($t(35.7) = -0.31$, $p = 0.76$).

1.3. Phonological short-term memory
Previous research has shown that phonological short-term memory (PSTM) predicts vocabulary learning (Baddeley, Papagno, & Vallar, 1988; Martin & Ellis, 2012). Given that learning multi-word units requires the ability to hold sequences longer than individual words in memory, PSTM was hypothesized to be of particular importance. A nonword repetition task was used to measure PSTM. Participants heard lists of one-syllable nonwords and repeated each list out loud. The lists employed here were adapted from Martin & Ellis (2012) to conform to Spanish phonotactics. Four lists of three, four, five or six nonwords were presented in ascending order (sample stimuli are provided below in Table 1). Participants' responses were recorded using a Zoom 4HN Pro digital recorder, and were scored following the criteria described in Gathercole, Pickering, Hall & Peaker. (2001, p. 15).

2. Measures of Cognitive control
To ensure that the two groups were comparable in terms of their cognitive control abilities, participants completed the AX-Continuous Performance Task (AX-CPT), the Flanker task, a Spanish working memory test.

2.1. AX-CPT
The AX-CPT has been used to measure individual styles of cognitive control. Participants saw sequences of five letters, the first and last of which were displayed in red font, and were asked to respond "yes" if the sequence started with the letter "A" and ended with an "X" (A – X). In all other conditions, they were instructed to respond "no."

The Dual Mechanisms of Control model (Braver, Gray & Burgess, 2007) considers the contributions of proactive and reactive control subcomponents. Because in some trials participants can anticipate a "no" response (B – X) while in others they

have to react rapidly (A – Y), an individual's Behavioral Shift Index (BSI) provides a relative measure of proactive/reactive control, with higher values indicating higher proactive control. Composite BSI scores were calculated following the methods described in Braver, Paxton, Locke and Barch (2009).

2.2. Flanker Task

The Flanker task provided an additional measure of executive control. A large body of studies have used the Flanker task to measure the association between language background and cognitive control (e.g., Emmorey, Luk, Pyers & Bialystok, 2008). In this task, participants responded to the direction of a central arrow in trials in which it is flanked by arrows pointing in the same direction (congruent trials), or the opposite direction (incongruent trials). The Flanker effect for each group was calculated by substracting reaction times in congruent trials from incongruent trials.

2.3. Reading Span Working Memory test

Participants were administered a Spanish version of the Reading Span Task, based on the Spanish adaptation of the original task in Daneman and Carpenter (1980) (Elosúa, Gutiérrez, García Madruga, Luque & Gárate, 1996). To ensure that sentences were processed as a whole, a manipulation was added in which participants were asked to judge whether each sentence made sense or not. As in Harrington and Sawyer's task (1992), half of the sentences were made ungrammatical by mixing up the order of the words.

Table 1. *Sample stimuli of nonword repetition task*

List length	3 items	4 items	5 items	6 items
	lib	dren	yir	cak
	chol	glach	gab	nej
	trum	nit	brok	yat
		lon	tep	mur
			chom	fram
				miz

Appendix B. List of stimuli

The table below presents the three lists of collocations that were studied and practiced by learners. Below each English collocation, its idiomatic Spanish translation is provided. Spanish meanings were presented during the Study phase, and in the recall tests. English collocations have literal Spanish equivalents, except for the verbs of incongruent collocations, for which the literal translation is not an adequate cross-linguistic equivalent. The literal English translations for the verbs of L1-L2 incongruent Spanish collocations are provided to the right of each verb.

	Incongruent collocations			Collocations with L1-related verb			Collocations with semantically-related verb		
	verb	(determiner)	noun	verb	(determiner)	noun	verb	(determiner)	noun
1	launder		money	whiten	one's	teeth	clean	one's	hands
	blanquear 'whiten'		*dinero*	*blanquear*		*dientes*	*limpiar*		*manos*
2	pack	one's	bags	make	a	cake	ready	the	room
	hacer 'make'		*maletas*	*hacer*		*tarta*	*preparar*		*habitación*
3	run	a	business	carry	one's	name	walk	a	street
	llevar 'carry'		*negocio*	*llevar*		*nombre*	*caminar*		*calle*
4	shoot	a	movie	roll	a	ball	fire	a	gun
	rodar 'roll'		*película*	*rodar*		*pelota*	*disparar*		*pistola*
5	file	a	complaint	put	an	end	arrange	a	meeting
	poner 'put'		*queja*	*poner*		*fin*	*organizar*		*reunión*
6	perform	a	song	touch	one's	hair	show		pictures
	tocar 'touch'		*canción*	*tocar*		*pelo*	*enseñar*		*imágenes*
7	raise		doubts	wake	one's	sister	increase	the	age

	English		Noun	Verb		Noun	Verb		Noun
	despertar 'wake'		*dudas*	*despertar*		*hermana*	*aumentar*		*edad*
8	meet	a	target	accomplish		work	join	one's	friends
	cumplir 'accomplish'		*objetivo*	*cumplir*		*trabajo*	*unirse*		*amigos*
9	take	a	walk	give		hope	catch	some	breath
	dar 'give'		*paseo*	*dar*		*esperanza*	*coger*		*aliento*
10	miss	one's	flight	lose		blood	find	the	truth
	perder 'lose'		*vuelo*	*perder*		*sangre*	*encontrar*		*verdad*
11	buy		time	win	the	war	pay	the	bill
	ganar 'win'		*tiempo*	*ganar*		*guerra*	*pagar*		*factura*
12	land	a	job	achieve	a	change	reach	a	height
	conseguir 'achieve'		*trabajo*	*conseguir*		*cambio*	*alcanzar*		*altura*
13	stuff	one's	mouth	fill	a	hole	load	a	truck
	llenar 'fill'		*boca*	*llenar*		*agujero*	*cargar*		*camión*
14	blow	the	bridge	fly	a	plane	break	one's	heart
	volar 'fly'		*puente*	*volar*		*avión*	*romper*		*corazón*
15	play	a	joke	spend	a	year	relax	the	mind
	gastar 'spend'		*broma*	*perder*		*año*	*relajar*		*mente*

Appendix C. Results of statistical analyses

1. Accuracy in recall tests

Table S1. *Model output for Recall Accuracy in Immediate and Delayed Tests*

	Estimate	Std. Error	|z|	Pr(>|z|)
(Intercept)	2.92	0.51	5.78	<0.0001
Condition Interference	0.8	0.45	1.78	0.07
Type Incong	-2.03	0.58	-3.49	<0.001
Type Semant	-0.34	0.59	-0.57	0.57
TestDelayed1	-0.62	0.23	-2.7	<0.01
TestImmediate2	1.66	0.3	5.5	<0.0001
TestDelayed2	1.66	0.3	5.5	<0.0001
TestDelayed3	0.66	0.26	2.57	<0.01
PSTM	0.71	0.21	3.4	<0.001
Condition Interference * Type Incong	0.81	0.2	4.03	<0.0001
Condition Interference * Type Semant	0.31	0.22	1.4	0.16
Type Incong * TestDelayed1	-0.92	0.26	-3.55	<0.001
Type Semant * TestDelayed1	-0.47	0.28	-1.65	0.1
Type Incong * TestImmediate2	-0.11	0.33	-0.34	0.73
Type Semant * TestImmediate2	0.12	0.38	0.31	0.76
Type Incong * TestDelayed2	-0.11	0.33	-0.34	0.73
Type Semant * TestDelayed2	0.12	0.38	0.31	0.76
Type Incong * TestDelayed3	-0.9	0.28	-3.16	<0.01
Type Semant * TestDelayed3	-0.15	0.32	-0.46	0.64
Condition Interference * TestDelayed1	-0.58	0.21	-2.76	<0.01
Condition Interference * TestImmediate2	-0.3	0.27	-1.09	0.27
Condition Interference * TestDelayed2	-0.3	0.27	-1.09	0.27
Condition Interference * TestDelayed3	-0.49	0.23	-2.18	<0.05

Dedicated Analyses on Delayed Tests 1 and 3				
Delayed Test 1				
(Intercept)	2	0.43	4.7	<0.001
Condition Interference	0.58	0.36	1.63	0.10
Type Incong	-2.4	0.48	-4.97	<0.001
Type Semant Rel	-0.86	0.49	-1.77	0.08
PSTM	0.57	0.18	3.15	<0.001
Delayed Test 3				
(Intercept)	3.74	0.65	5.74	<0.001
Condition Interference	0.47	0.56	0.83	0.41

Type Incong	-3.14	0.75	-4.2	<0.001
Type Semant Rel	-0.62	0.75	-0.82	0.41
PSTM	0.86	0.25	3.42	<0.001
Condition Interf * Type Incong	0.8	0.44	1.83	0.07
Condition Interf * Type Semant	0.34	0.48	0.71	0.48

2. Growth Curve Analysis of Reaction Time for verb selection in incongruent collocations

Table S2. *GCA output for RTs in incongruent collocations*

	Estimate	Std. Error	$|t|$	$Pr(>|t|)$
Practice session 1				
(Intercept)	-0.04	0.06	-0.59	0.56
ot1	-0.88	0.29	-2.98	<0.01
ot2	0.64	0.22	2.94	<0.01
Condition Interference	-0.06	0.05	-1.07	0.29
PSTM	-0.05	0.02	-2.90	<0.01
ot1 * Condition Interference	-1.09	0.42	-2.59	<0.01
Practice Session 2				
(Intercept)	-0.15	0.05	-3.12	<0.01
ot1	-0.71	0.27	-2.63	<0.01
ot2	0.43	0.17	2.47	<0.05
Practice Session 3				
(Intercept)	-0.07	0.06	-1.07	0.28
ot1	-0.38	0.22	-1.73	0.08
ot2	0.27	0.17	1.6	0.11
Condition_Interference	-0.1	0.04	-2.23	<0.05

3. Analysis of RT for verb selection across collocation types

Table S3. *Results of the RT mixed-effects regression analysis of verb selection for collocation types*.

	Estimate	Std. Error	$|t|$	$Pr(>|t|)$
Practice session 1				
(Intercept)	0.27	0.11	2.52	0.01
PSTM	-0.01	0.08	-0.13	0.9
Type Semantic	0.01	0.09	0.08	0.94
Type Incongruent	-0.02	0.09	-0.22	0.83
Flanker Effect	-0.09	0.08	-1.1	0.27
PSTM * Type Semantic	-0.04	0.02	-2.14	<0.05
PSTM * Type Incongruent	-0.06	0.02	-3.06	<0.01
Type Semantic * Flanker_Effect	-0.05	0.02	-2.54	<0.01
Type Incongruent * Flanker Effect	-0.06	0.02	-2.75	<0.01

Practice session 2

(Intercept)	0.04	0.09	0.49	0.62
PSTM	-0.02	0.06	-0.34	0.73
Type Semantic	-0.03	0.09	-0.36	0.72
Type Incongruent	-0.1	0.09	-1.1	0.27
PSTM * Type Semantic	-0.04	0.02	-2.51	<0.05
PSTM * Type Incongruent	-0.06	0.02	-3.47	<0.001

Practice session 3

(Intercept)	-0.17	0.1	-1.8	0.07
Type Semantic	-0.01	0.07	-0.14	0.9
Type Incongruent	-0.01	0.07	-0.15	0.88
Condition Interference	0.01	0.12	0.06	0.95
PSTM	-0.02	0.06	-0.34	0.73
Condition Interference * Type Semantic	-0.01	0.03	-0.44	0.66
Condition Interference * Type Incongruent	-0.11	0.03	-3.34	<0.001
PSTM * Type Semantic	-0.04	0.02	-2.65	<0.01
PSTM * Type Incongruent	-0.04	0.02	-2.72	<0.01

References

Baddeley, A., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language, 27*, 586-595. doi:10.1016/0749-596X(88)90028-9

Braver, T. S., Gray, J. R., & Burgess, G. C. (2008). *Explaining the many varieties of working memory variation: Dual mechanisms of cognitive control*. New York: Oxford University Press. doi:10.1093/acprof:oso/9780195168648.003.0004

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences, 106*, 7351. doi:10.1073/pnas.0808187106

Elosúa, M. R., Gutiérrez, F., Juan Antonio García Madruga, Luque, J. L., & Gárate, M. (1996). Adaptación Española del "reading span" test de Daneman y Carpenter. *Psicothema, 8*

Emmorey, K., Luk, G., Pyers, J., & Bialystok, E. (2008). The source of enhanced cognitive control in bilinguals: Evidence from bimodal bilinguals. *Psychological Science, 19*, 1201–1206.

Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *The Quarterly Journal of Experimental Psychology Section A, 54*, 1. doi:10.1080/02724980042000002

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research, 50*, 940-967. doi:10.1044/1092-4388(2007/067)

Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition, 34*, 379. doi:10.1017/S0272263112000125