Supplementary Online Materials (SOM) to *Changing expectations mediate adaptation in L2 production*

Guillermo Montero-Melis[1,2,3], T. Florian Jaeger[4,5,6]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2] Department of Linguistics, Stockholm University, Stockholm, Sweden
[3] Centre for Research on Bilingualism, Stockholm University, Stockholm Sweden
[4] Department of Brain and Cognitive Sciences, University of Rochester
[5] Department of Computer Science, University of Rochester
[6] Center for Language Sciences, University of Rochester

Contents

Appendix S1. Stimuli Description and Priming Sentences

Table S1. Description of target stimuli and priming sentences used in the primed conditions.

#	Path	Manner	Direction	Object	Ground	Manner-priming sentence[1]	Path-priming sentence
1	down	push + slide	right-left	suitcase	hill	El hombre empuja un sofá por una rampa. (‘The man pushes a sofa along a ramp.’)	El hombre baja una rampa con un sofá. (‘The man descends a ramp with a sofa.’)
2	down	push + slide	left-right	suitcase	snowy hill	El chico empuja una piedra por una pendiente. (‘The boy pushes a stone along a slope.’)	El chico baja una pendiente con una piedra. (‘The boy descends a slope with a stone.’)
3	into	push + slide	right-left	table	cave	El empleado empuja un armario por un almacén. (‘The employee pushes a cupboard in a warehouse.’)	El empleado entra en un almacén con un armario. (‘The employee enters a ware-house with a cupboard.’)
4	into	push + slide	left-right	table	house	El informático empuja un orde-nador por un cuarto. (‘The computer technician pushes a computer in a room.’)	El informático entra en un cuarto con un ordenador. (‘The computer technician enters a room with a com-puter.’)
5	up	push + slide	left-right	package	dune	El señor empuja una televisión por unos escalones. (‘The man pushes a television along some steps.’)	El señor sube unos escalones con una televisión. (‘The man ascends some steps with a television.’)
6	up	push + slide	right-left	package	roof	La moza empuja un sillón por un repecho. (‘The young girl pushes an arm-chair along a slope.’)	La moza sube un repecho con un sillón. (‘The young girl ascends a slope with an armchair.’)
7	across	push + slide	left-right	basket	road	El niño empuja una papelera por una habitación. (‘The boy pushes a basket in a bedroom.’)	El niño cruza una habitación con una papelera. (‘The boy crosses a bedroom with a basket.’)
8	across	push + slide	right-left	basket	street	El bombero empuja un piano por un pasillo. (‘The fireman pushes a piano along a hallway.’)	El bombero cruza un pasillo con un piano. (‘The fireman crosses a hallway with a piano.’)
9	down	push + roll	right-left	balloon	hill	La niña rueda un plato por una cuesta. (‘The girl rolls a plate along a slope.’)	La niña baja una cuesta con un plato. (‘The girl descends a slope with a plate.’)
10	down	push + roll	left-right	balloon	snowy hill	El camarero rueda un barril por unos peldaños. (‘The waiter rolls a keg along some steps.’)	El camarero baja unos peldaños con un barril. (‘The waiter descends some steps with a keg.’)
11	into	push + roll	left-right	tire	cave	La joven rueda un aro por un gimnasio. (‘The girl rolls a hoop in a gym.’)	La joven entra en un gimnasio con un aro. (‘The girl enters a gym with a hoop.’)
12	into	push + roll	right-left	tire	house	La chiquilla rueda un hula hoop por un aula. (‘The little girl rolls a hula hoop in a classroom.’)	La chiquilla entra en un aula con un hula hoop. (‘The little girl enters a classroom with a hula hoop.’)

[1] Note that the Spanish preposition *por* is used in all manner-priming sentences. This preposition is semantically underspecified compared to typical English path prepositions and satellites; it can denote motion inside a location (e.g., item 3) as well as motion along a landmark (e.g., item 1). English translations are difficult to render using the same preposition for all items. We have therefore opted for alternating between different translations of *por* to make the sentences more intelligible, even though we are fully aware that the English translations are not idiomatic, precisely because both path and manner information are typically expressed in English.

#	Path	Manner	Direction	Object	Ground	Manner-priming sentence[1]	Path-priming sentence
13	up	push + roll	left-right	swimming ring	dune	El escarabajo rueda una pelota por un montículo. ('The beetle rolls a ball along a mound.')	El escarabajo sube un montículo con una pelota. ('The beetle ascends a mound with a ball.')
14	up	push + roll	right-left	swimming ring	roof	El leñador rueda un tronco por una loma. ('The lumberjack rolls a log along a hill.')	El leñador sube una loma con un tronco. ('The lumberjack ascends a hill with a log.')
15	across	push + roll	left-right	cartwheel	road	El mozo rueda una esfera por una plaza. ('The young boy rolls a sphere/ball on a square.')	El mozo cruza una plaza con una esfera. ('The young boy crosses a square with a sphere/ball.')
16	across	push + roll	right-left	cartwheel	street	El borracho rueda una botella por una avenida. ('The drunkard rolls a bottle along an avenue.')	El borracho cruza una avenida con una botella. ('The drunkard crosses an avenue with a bottle.')
17	down	pull + slide	left-right	trunk	hill	El conserje arrastra un colchón por una escalinata. ('The concierge drags a mattress along a staircase.')	El conserje baja una escalinata con un colchón. ('The concierge descends a staircase with a mattress.')
18	down	pull + slide	right-left	trunk	snowy hill	El obrero arrastra un tablón por unas gradas. ('The labourer drags a board along a grandstand.')	El obrero baja unas gradas con un tablón. ('The labourer descends a grandstand with a board.')
19	into	pull + slide	right-left	chair	cave	La mujer arrastra una cómoda por un vestíbulo. ('The woman drags a chest in a hall.')	La mujer entra en un vestíbulo con una cómoda. ('The woman enters a hall with a chest.')
20	into	pull + slide	left-right	chair	house	El vendedor arrastra una lámpara por un local. ('The sales agent drags a lamp in a store.')	El vendedor entra en un local con una lámpara. ('The sales enters a store with a lamp.')
21	up	pull + slide	left-right	sack	dune	La hormiga arrastra una hoja por una roca. ('The ant drags a leaf along a rock.')	La hormiga sube una roca con una hoja. ('The ant ascends a rock with a leaf.')
22	up	pull + slide	right-left	sack	roof	El explorador arrastra un trineo por un desnivel. ('The explorer drags a sleigh along a slope.')	El explorador sube un desnivel con un trineo. ('The explorer ascends a slope with a sleigh.')
23	across	pull + slide	right-left	rocking horse	road	La anciana arrastra una alfombra por una sala. ('The old lady drags a rug in a living room.')	La anciana cruza una sala con una alfombra. ('The old lady crosses a living room with a rug.')
24	across	pull + slide	left-right	rocking horse	street	El campesino arrastra un palo por una parcela. ('The peasant drags a stick along a plot.')	El campesino cruza una parcela con un palo. ('The peasant crosses a plot with a stick.')
25	down	pull + roll	left-right	wheel-barrow	hill	La muchacha tira de una bicicleta por una costana. ('The young girl draws a bicycle along a steep road.')	La muchacha baja una costana con una bicicleta. ('The young girl descends a steep road with a bicycle.')
26	down	pull + roll	right-left	wheel-barrow	snowy hill	El muchacho tira de una moto por un cerro. ('The young boy draws a motorbike along a hill.')	El muchacho baja un cerro con una moto. ('The young boy descends a hill with a motorbike.')
27	into	pull + roll	left-right	shopping trolley	cave	El chiquillo tira de un patinete por un jardín.	El chiquillo entra en un jardín con un patinete.

#	Path	Manner	Direction	Object	Ground	Manner-priming sentence[1]	Path-priming sentence
						(‘The kid draws a scooter in a garden.’)	(‘The kid enters a garden with a scooter.’)
28	into	pull + roll	right-left	shopping trolley	house	El nene tira de un triciclo por una escuela. (‘The little child draws a tricycle in a school.’)	El nene entra en una escuela con un triciclo. (‘The little child enters a school with a tricycle.’)
29	up	pull + roll	left-right	toy car	dune	La pasajera tira de un bolso con ruedas por una pasarela. (‘The passenger draws a rolling bag along a gangway.’)	La pasajera sube una pasarela con un bolso con ruedas. (‘The passenger ascends a gangway with a rolling bag.’)
30	up	pull + roll	right-left	toy car	roof	La labradora tira de un carro por un collado. (‘The farmer draws a cart along a hillock.’)	La labradora sube un collado con un carro. (‘The farmer ascends a hillock with a cart.’)
31	across	pull + roll	right-left	stroller	road	El labrador tira de una carreta por un puente. (‘The farmer draws a wagon along a bridge.’)	El labrador cruza un puente con una carreta. (‘The farmer crosses a bridge with a wagon.’)
32	across	pull + roll	left-right	stroller	street	El joven tira de un monopatín por una vía. (‘The young man draws a skateboard along a road.’)	El joven cruza una vía con un monopatín. (‘The young man crosses a road with a skateboard.’)

Appendix S2. Participant background information

Because individual differences and demographic information is of potential interest in the field of second language acquisition and bilingualism, and more generally as part of good scientific practice, we here report all background information collected about the participants.[2] We distinguish between measures directly or indirectly related to L2 proficiency and other background information collected as part of a standard data collection procedure. Only the former are of potential theoretical relevance, and for this reason we did not investigate the effect of any other variables on the dependent variables considered in this study. Among the L2 proficiency measures, we only ever considered the outcome of the Cloze test as a predictor for our analyses, as we expected this to be the most sensitive and objective of all proficiency measures (Tremblay, 2011), and in order to avoid inflation of Type I error rates due to unnecessary researchers' degrees of freedom (Simmons, Nelson, & Simonsohn, 2011). We also note that we used random assignment to assign participants to the three different between-subject conditions (path primed, manner primed, baseline), which is the preferred approach instead of trying to account for the effect of covariates or trying to create balance by assigning participants to experimental conditions depending on a covariate (Vanhove, 2015).

Measures related to L2 proficiency

The main L2 proficiency measure we collected and the only one we used as a predictor in the analyses was the score obtained on a cloze test (see main document Method > Participants). The distribution of cloze scores by condition is shown in Figure S2-1. We additionally collected the following measures: 1) self-ratings of participants' proficiency in Spanish (*L2_SelfRating*), which were collected at the recruitment stage, before the experimental session, and could range on a scale from 1 (very low proficiency) to 7 (very

[2] We are grateful to an anonymous reviewer for this request.

high proficiency); 2) self-reported age of onset (*L2_AoO*) at which a participant first started to learn or came into contact with Spanish; 3) self-reported years of formal instruction in Spanish (*L2_instruction_years*); 4) accuracy on a vocabulary task (*VocabTaskAcc*) which was deliberately easy (see main document Method > Procedure). Pairwise scatterplots and Spearman's rank correlation coefficients between all variables potentially related to L2 proficiency, as well as density plots for each variable are shown in Figure S2-2. None of these measures differed significantly between priming condition (all Kruskal-Wallis tests $p >$ .5).

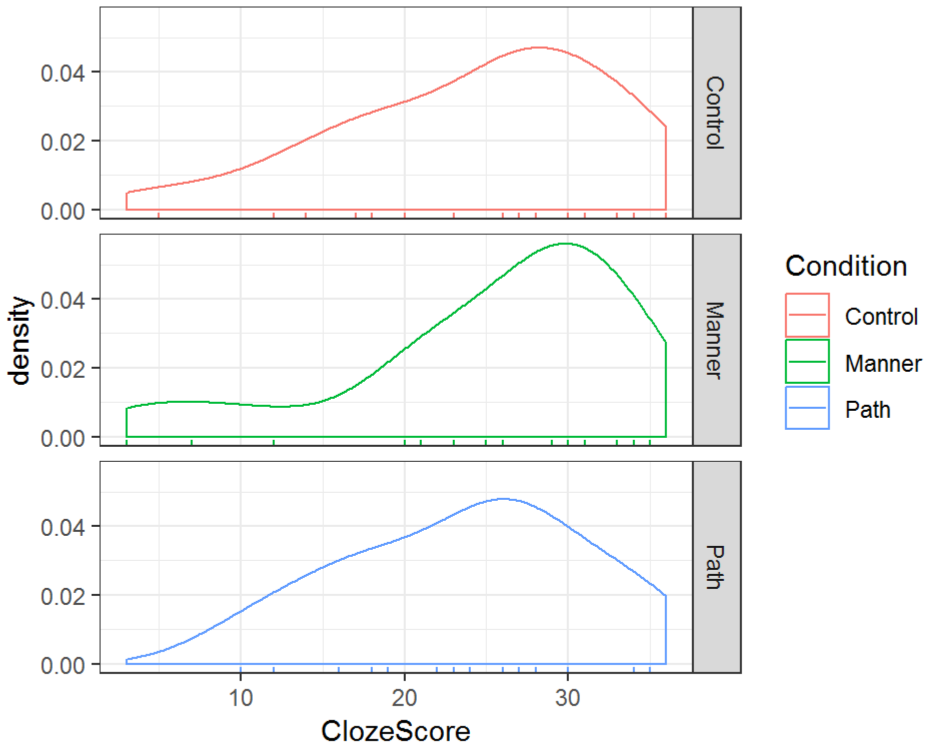

Figure S2-1. Distribution of Cloze scores across conditions. Density curves with individual observations plotted as a rug plot along the x-axis.

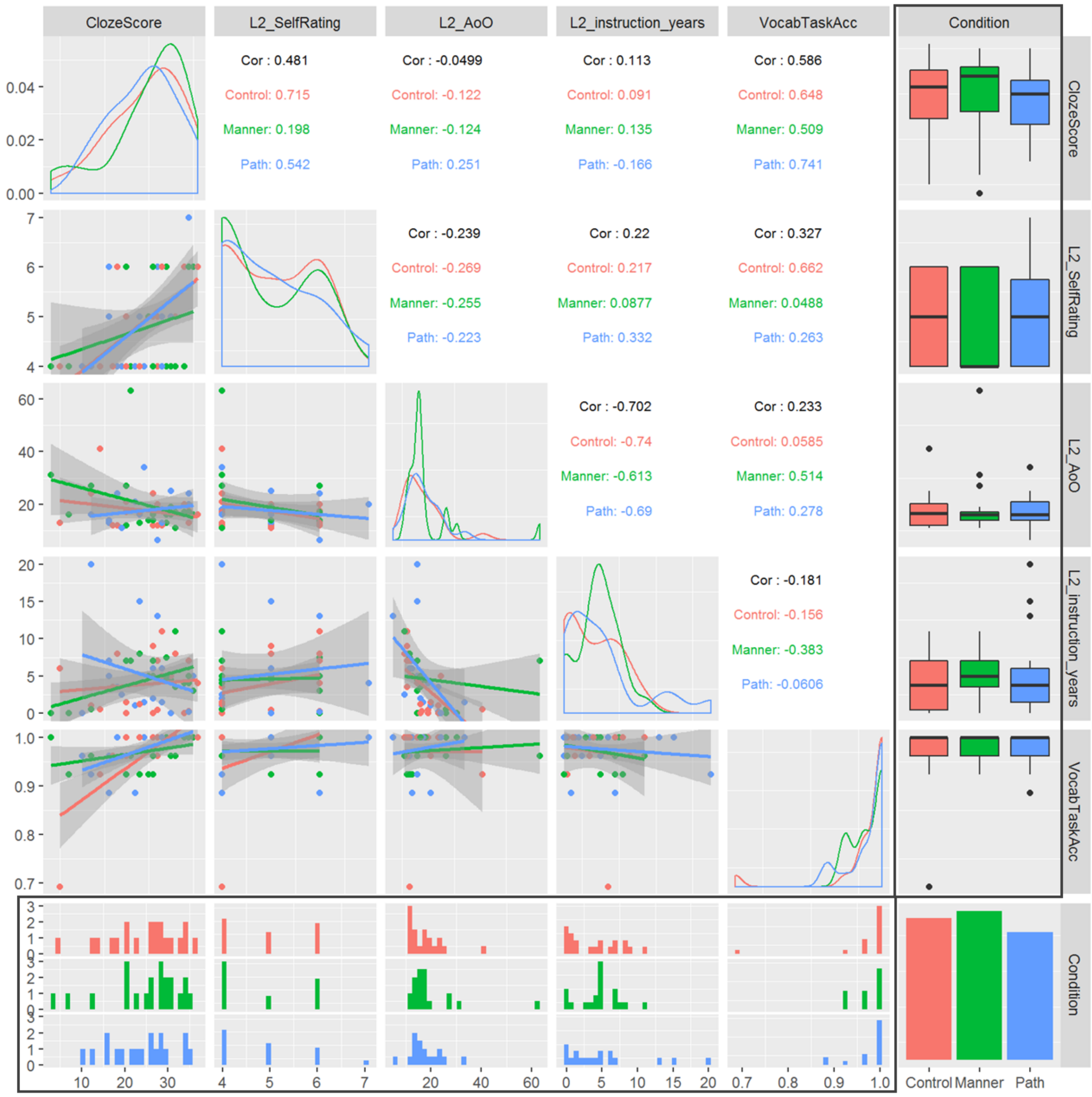

Figure S2-2. Pairwise scatterplots (lower triangular) and Spearman's rank correlation coefficients (upper triangular) between all variables potentially related to L2 proficiency, coloured by condition. Plots on the diagonal show density plots for each variable by condition. For each variable, the bottom row shows histograms by condition, and the rightmost column shows boxplots by condition (with the horizontal line corresponding to the median.

Other background information

We collected information about the gender and age of the participants. Gender distribution among groups is shown in Table S2-1. Statistical summaries of age are given in the main document (see Method > Participants); density curves and rug plots for this variable broken down by group and condition are shown in Figure S2-3. Neither of these variables was used in the analyses.

Table S2-1. Distribution of gender across groups.

Group	#females	#males
Native Spanish speakers	27	32
L2 speakers	40	19

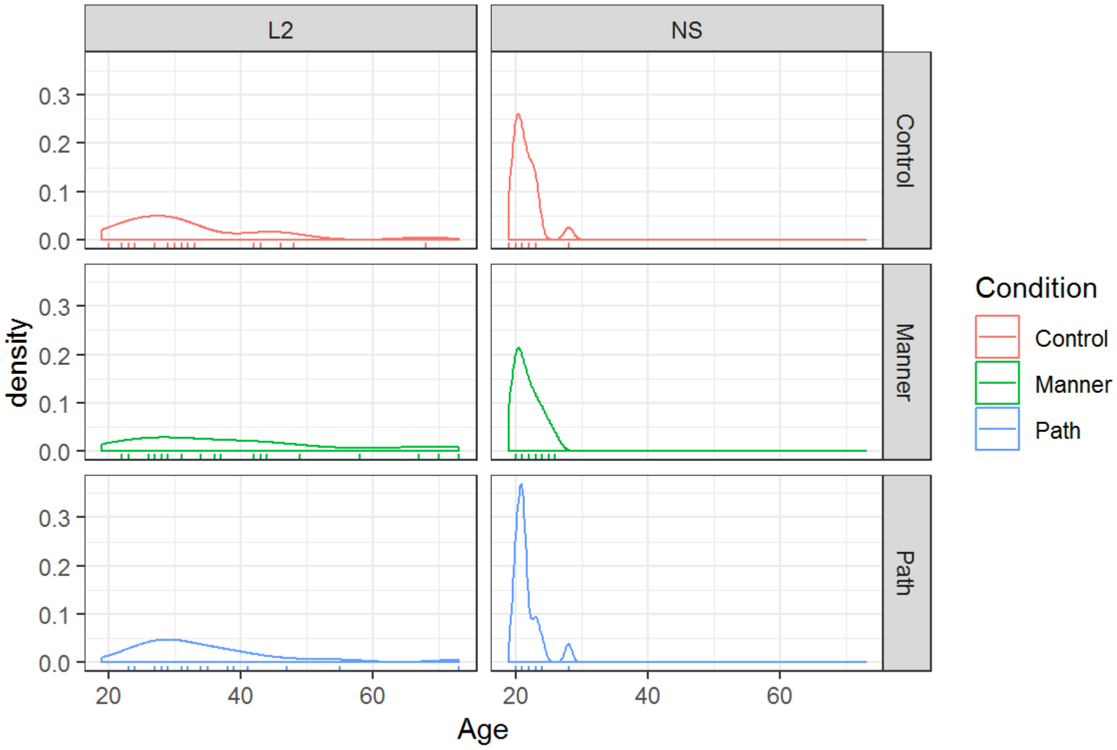

Figure S2-3. Distribution of age across group and conditions. Each panel shows density curves with individual observations plotted as a rug plot along the x-axis.

Appendix S3. GLMM fitted to data from baseline conditions

The complete model output fitted to the data from the baseline conditions (see main document Results > Baseline condition) is shown in Figure S3.

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: Used ~ VerbType * Group + (1 + VerbType | Subject) + (1 + VerbType *      Group | VideoName)
   Data: d_basel
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))

     AIC      BIC   logLik deviance df.resid 
  2378.3   2477.7  -1172.2   2344.3     2535 

Scaled residuals: 
    Min      1Q  Median      3Q     Max 
-6.3134 -0.4718 -0.1630  0.4942  4.8860 

Random effects:
 Groups    Name                         Variance Std.Dev. Corr             
 Subject   (Intercept)                  0.597562 0.77302                   
           VerbTypeM_vs_P               3.100589 1.76085  -0.15            
 VideoName (Intercept)                  0.021985 0.14827                   
           VerbTypeM_vs_P               0.347587 0.58956   0.18            
           GroupL2_vs_NS                0.007223 0.08499   0.71  0.82      
           VerbTypeM_vs_P:GroupL2_vs_NS 0.124765 0.35322   0.90 -0.26  0.33
Number of obs: 2552, groups:  Subject, 40; VideoName, 32

Fixed effects:
                             Estimate Std. Error z value Pr(>|z|)    
(Intercept)                   -0.6586     0.1458  -4.516 6.31e-06 ***
VerbTypeM_vs_P                -1.2260     0.3073  -3.989 6.63e-05 ***
GroupL2_vs_NS                 -0.3113     0.1413  -2.204   0.0276 *  
VerbTypeM_vs_P:GroupL2_vs_NS  -0.4233     0.2945  -1.437   0.1506    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) VrTM__P GL2__N
VrbTypM_v_P -0.064               
GrpL2_vs_NS  0.065  0.051        
VTM__P:GL2_  0.055 -0.006  -0.091
```

Figure S3. Complete output for the generalized linear mixed model fitted to the data from the baseline conditions (as obtained from the *summary* function in R).

Appendix S4. GAMM fitted to native speaker data (Question 1) with glossed model output and follow-up analysis

Glossed GAMM output

Some aspects of Generalized Additive Mixed Model (GAMM) analyses differ from more traditional analytical approaches like ANOVAs or generalized linear regression. For example, the estimated coefficients in GAMMs are not directly informative of the shape or even direction of an effect. Therefore model visualization becomes essential (Winter & Wieling, 2016). Other aspects of GAMMs are familiar from Generalized Linear Mixed Models (GLMMs; for introduction, see Baayen, Davidson, & Bates, 2008). First, like GLMMs, methods for estimating statistical significance (i.e., *p*-values) are approximate and still an area of development in GAMMs (e.g., Wood, 2013). Another commonality with GLMMs is that GAMMs, too, can model random variation introduced by subjects or items. In addition to random intercepts or random slopes, however, GAMMs allow for so-called smooth factors, which conceptually function much like random intercepts and slopes in GLMMs, except that smooth factors also account for random variation by units (e.g., participants) in the *functional shape* of the relation between predictor variables and outcome variables, including markedly non-linear shapes that may differ by subject (for practical illustration, see van Rij, 2015).

For the GAMM fitted to native speaker data (see 'Trial-by-trial adaptation of native speakers (Question 1)' in main document), we first explain the model specification in the function call, and then gloss the model output in some detail so as to facilitate comprehension and familiarize the reader with the meaning of the different coefficients.

The function call and model specification in R for this GAMM was:[3]

---

[3] We initially entertained a more complex random effect structure, including by-item factor smooths for Trial. However, this level of complexity was not justified by the data and so we simplified random effects by items to a by-item random intercept. None of the relevant coefficients reported here and none of the conclusions are affected by this choice. Please see the analysis script in the Dataverse repository for the details (at https://doi.org/10.7910/DVN/TOJ1UH).

```
mgcv::bam(Used ~ VbType_Cond + s(Trial, by = VbType_Cond) + s(Trial,
Subject, bs = 'fs') + s(VideoName, bs = 're'), data = d_ns, family =
'binomial')
```

In the model formula, ‘Used’ represents the binary outcome (1 vs. 0). ‘VbType_Cond’ is the four-level ‘Verb Type-by-Condition’ factor obtained by crossing the levels of Condition (primed vs. baseline) with those of the indicator variable Verb Type (manner vs. path verb). Recall from the main document: “When the value of the indicator variable is ‘path’, then an outcome of 1 means that the main verb in a participant’s description on that trial expressed path information, and an outcome of 0 means the main verb did not express path. When the value of the indicator variable is ‘manner’, then an outcome of 1 means that the main verb expressed manner, and an outcome of 0 means it did not” (p. 15). The predictor ‘Trial’ indicates the order in which scenes were described. The syntax ‘s(Trial, by = VbType_Cond)’ indicates that ‘Trial’ was included in the model as a potentially non-linear smooth that could interact with Condition and Verb Type (i.e., the smooth could differ for each of the four levels of ‘VbType_Cond’). The term ‘s(Trial, Subject, bs = 'fs')’ indicates that the model included random by-participant factor smooths for Trial to capture variability between

Table S4. Summary of the GAMM for trial-by-trial adaptation in native speakers. The explained deviance of the model is 38.6%.

A. parametric coefficients	**Estimate**	**Std. Error**	**z-value**	**p-value**
Intercept (Path-verb_Baseline)	0.3980	0.4080	0.9755	0.3293
Path-verb_Path-priming	0.9573	0.5977	1.6017	0.1092
Manner-verb_Baseline	-1.4606	0.5819	-2.5101	0.0121
Manner-verb_Manner-priming	0.0490	0.5702	0.0860	0.9315
B. smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
s(Trial):Path-verb_Baseline	1.0000	1.0000	9.3006	0.0023
s(Trial):Path-verb_Path-priming	1.5965	1.9777	13.9674	0.0013
s(Trial):Manner-verb_Baseline	1.0000	1.0000	10.0372	0.0015
s(Trial):Manner-verb_Manner-priming	2.0599	2.5715	5.0871	0.1374
s(Trial,Subject)	103.0745	782.0000	641.5973	< 0.0001
s(Item)	0.0002	31.0000	0.0001	0.5454

speakers. The final term ‘s(VideoName, bs = 're')’ indicates that the model included random by-item intercepts (the variable ‘VideoName’ identifies each of the 32 items).

The summary of the GAMM for trial-by-trial adaptation in native speakers is shown in Table S4 (and visualized in Figure 4 in the main document). Table S4 is divided into parametric coefficients and smooth terms (table sections A and B, respectively). The former (parametric coefficients) can be interpreted in the same way as fixed effect coefficients in GLMMs, as follows. The intercept term in this and in all GAMMs reported below refers to the log-odds of a path verb in the baseline condition. Here the estimates show that participants in the baseline condition were somewhat more likely to use path main verbs than not to use them, but the coefficient is not significantly different from a log-odds of zero (row 1: estimate = 0.40, z = 0.98, p = .33). Note that path verbs in the baseline condition (the intercept) is the reference level with respect to which the other parametric coefficients are estimated (using simple or “dummy” coding). The subsequent rows in section A of Table S4 respectively indicate that participants in the path-primed condition (“Path-verb_Path-priming”) showed a non-significant trend towards using more path verbs than participants in the baseline condition (i.e., the reference level) (row 2: estimate = 0.96, z = 1.60, p = .11); that the log-odds of using a manner verb in the baseline condition (“Manner-verb_Baseline”) were significantly lower than the log-odds of using a path verb in the baseline condition (row 3: estimate = −1.46, z = −2.51, p = .01); and, finally, that the log-odds of using a manner verb in the manner-primed condition (“Manner-verb_Manner-priming”) did not significantly differ from the log-odds of using a path verb in the baseline condition (row 4: estimate = 0.05, z = 0.09, p > .9).

The part of the model summary that is specific to GAMMs is shown in section B of Table S4 (smooth terms). Interpretation is facilitated by checking the visualization in Figure 4 (main document) in parallel. Consider the first row, “s(Trial):Path-verb_Baseline,” which refers to the model prediction of the use of path verbs in the baseline condition as a smooth (possibly nonlinear) function of trial. The first column reports the effective degrees of freedom (*edf*), which is a measure of the complexity of the smooth function: When there is a single smooth predictor (as here, Trial), then *edf*s close to 1 indicate that the estimated functional form is close to a straight line, whereas values greater than one indicate the smooth functions are increasingly complex (i.e., more nonlinear and wiggly; see Wood, 2006).[4] The low p-value (p = .002) for “s(Trial):Path-verb_Baseline” indicates that this approximately straight line has a slope significantly different from zero. This can be seen in Figure 4 in the main document (dashed blue line in upper left panel): this line—which provides the best trade-off between complexity and fit to the data—is close to a straight line (hence $edf \approx 1$), but it is not horizontal nor can a horizontal line easily be fit within the shaded area representing the 95% confidence interval (hence p = .002). This means that native speakers in the baseline condition showed a tendency to use more path verbs as the experiment proceeded.

The second row of Table S4-B, “s(Trial):Path-verb_Path-priming,” is interpreted in an analogous fashion: the slightly higher *edf* value of 1.6 indicates a more wiggly (i.e., less linear) function of trial for path-verb use under path-verb priming. This is visually corroborated by the corresponding slightly convex curve in Figure 4 in the main document (see red continuous line in upper left panel).

Row 3 in Table S4-B, “s(Trial):Manner-verb_Baseline,” indicates a linear relation between trial and the log-odds of manner verbs in the baseline condition (edf = 1), whose slope nevertheless is different from zero (p = .002), similar to what was observed in row 1.

---

[4] Critically, whereas linearity is an *assumption* in GLMMs, a linear or close-to-linear relation (i.e., *edfs* close to 1) in the GAMM framework is a *finding*. The *edfs* that the fitting procedure underlying the GAMM assigns to a non-parametric predictor reflect a balance of fit against the data versus complexity of the model (measured in terms of *edf*s). Another way to think about this is that the *edf*s are a result of trying to come up with a model that provides a good fit against the data, but also is likely to generalize to (i.e., be predictive about) unseen data, and thus avoids over-fitting the data.

This again is corroborated by the blue dashed line in the upper right panel of Figure 4 in the main document, which shows that baseline participants—who tended to use more path verbs as the experiment proceeded—also tended to use less manner verbs as the experiment progressed. Comparison of rows 1 and 3 of Table S4-B highlights an important point concerning GAMMs: the output does not contain information about the direction of the effects. Here, the numerical information from the model output is almost identical for these two rows. It is only by checking the model estimates visually (e.g., Figure 4 in main document) that the direction of effects becomes clear.

The fourth row in Table S4-B indicates that the effect of trial on manner-verb use under manner-verb priming was non-linear (*edf* = 2.1), and yet it did not differ significantly from a flat horizontal line ($p = .14$). This corresponds well to the red line in the upper right panel of Figure 4 in the main document.

Finally, rows 5 and 6 in Table S4-B correspond to the GAMM analogue of the random effects in a GLMM. The high *edf* and low *p*-value for the factor smooth for subjects ("s(Trial, Subject)") suggests that this term captured much idiosyncratic participant variability in the data, whereas the low values for "s(Item)" (i.e., random by-item intercepts) suggests that there was little random variation associated with the specific events in the material.

Follow-up analysis

Figure 4 in the main document shows that native speakers significantly adapted to manner verbs, but not to path verbs. This, however, does not necessarily entail that their patterns of adaptation to path and manner verbs differed significantly. Unfortunately, the GAMM does not currently allow us to numerically compare the strength of adaptation in each condition (confirmed with Simon Wood, p.c.), which is a question about how quickly speakers adapt to one or the other lexicalization pattern. A suboptimal post-hoc approach to compare whether there was a difference in the strength of adaptation to path or manner verbs among native speakers is to model the data using GLMMs. This class of models is appropriate to test linear relations (here, in log-odds space) between predictors and an outcome variable, and allows for statistical comparison of the slopes of interest (i.e., the slopes of the linear part of the curves in the lower panels of Figure 4).

This approach necessarily suffers from power loss. We may lose power either a) because by the GLMM assumptions, we are modelling as linear (in log-odds space) an effect that is non-linear, thus losing sensitivity to capture the relevant signal, or b) because we reduce the number of observations to include only data points that seem to have a linear relationship between predictors and dependent variable—or due to a combination of the two.

With the above caveats in mind, we ran a GLMM analysis by subsetting the data to use only those trials that show an approximately linear relation with the outcome variable (based on visual inspection), keeping trials 15 through 32. The model included as fixed effect predictors: Priming Condition (sum-coded: 1 = primed, −1 = baseline), Verb Type (sum-coded: 1 = manner verb, −1 = path verb), and Trial (trials 15 through 32, centred around the mean), as well as all their interactions. The random effect structure consisted of random by-subject and by-item intercepts, and a random by-subject slope for the centred Trial predictor, which was the random effect structure that was conceptually analogous to that used in the GAMM. The function call and model specification in R was:

```
lme4::glmer(Used ~ Condition * VerbType * cTrial + (1 + cTrial |
Subject) + (1 | VideoName), data = d_ns_linear, family = 'binomial',
control = glmerControl(optimizer = 'bobyqa', optCtrl =
list(maxfun=2e5)))
```

The complete model output is shown in Figure S4.The critical fixed-effect coefficient in the model to assess if the strength of adaptation differed between priming conditions is the three-way interaction between Priming Condition, Verb Type and (centred) Trial. This coefficient estimates whether the slope on trials 15–32 in the lower left panel of Figure 4

‘Path-primed – Baseline’ is steeper than the corresponding slope in the lower right panel of Figure 4 ‘Manner-primed – Baseline’ (see main document). The coefficient was numerically positive, indicating that, as predicted, the slope was estimated to be steeper for manner-verb adaptation than for path-verb adaptation, but the coefficient was non-significant (estimate = 0.02, z = 1.04, p = .30).

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: Used ~ Condition_bin * VerbType * cTrial + (1 + cTrial | Subject) +      (1 | VideoName)
   Data: d_ns_lin
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))

     AIC      BIC   logLik deviance df.resid 
  1338.6   1401.7   -657.3   1314.6     1408 

Scaled residuals: 
    Min      1Q  Median      3Q     Max 
-3.5388 -0.4893  0.1218  0.4862  3.2109 

Random effects:
 Groups    Name        Variance Std.Dev. Corr
 Subject   (Intercept) 5.285923 2.29911      
           cTrial      0.004483 0.06696  0.89
 VideoName (Intercept) 0.000000 0.00000      
Number of obs: 1420, groups:  Subject, 79; VideoName, 32

Fixed effects:
                                                      Estimate Std. Error z value Pr(>|z|)    
(Intercept)                                            0.44254    0.27849   1.589 0.112042    
Condition_binPrimed_vs_Baseline                        0.87788    0.28038   3.131 0.001742 ** 
VerbTypeM_vs_P                                        -1.14788    0.28170  -4.075  4.6e-05 ***
cTrial                                                 0.03211    0.01680   1.911 0.056023 .  
Condition_binPrimed_vs_Baseline:VerbTypeM_vs_P         0.25768    0.27844   0.925 0.354730    
Condition_binPrimed_vs_Baseline:cTrial                 0.04387    0.01794   2.446 0.014453 *  
VerbTypeM_vs_P:cTrial                                 -0.06337    0.01835  -3.454 0.000552 ***
Condition_binPrimed_vs_Baseline:VerbTypeM_vs_P:cTrial  0.01742    0.01673   1.041 0.297815    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
               (Intr) Cn_P__B VrTM__P cTrial Cn_P__B:VTM__P C_P__B:T VTM__P:
Cndtn_bP__B     0.027                                                       
VrbTypM_v_P    -0.048 -0.084                                                
cTrial          0.461  0.028  -0.046                                        
Cn_P__B:VTM__P -0.059 -0.030   0.008  -0.037                                
Cndt_P__B:T     0.029  0.452  -0.083   0.094 -0.008                         
VrbTyM__P:T    -0.043 -0.079   0.458  -0.124 -0.006         -0.276          
C_P__B:VTM__P: -0.037 -0.009  -0.008  -0.089  0.460          0.030   -0.057 
```

Figure S4. Follow-up analysis to Question 1. Complete output for the generalized linear mixed model fitted to the native speaker data (as obtained from the *summary* function in R).

Appendix S5. GAMM fitted to L2 learners as a group (Question 2) and follow-up analysis

GAMM output

The GAMM fitted to the L2 learner data (see 'Trial-by-trial adaptation of L2 learners

Table S5. Summary of the GAMM for trial-by-trial adaptation in L2 learners. The explained deviance of the model is 54.9%.

A. parametric coefficients	**Estimate**	**Std. Error**	**z-value**	**p-value**
Intercept (Path-verb_Baseline)	0.5949	0.4116	1.4454	0.1484
Path-verb_Path-priming	2.4372	0.6270	3.8872	0.0001
Manner-verb_Baseline	-2.9603	0.5941	-4.9829	< 0.0001
Manner-verb_Manner-priming	-0.1649	0.5765	-0.2860	0.7749
B. smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
s(Trial):Path-verb_Baseline	2.5646	3.1916	7.5895	0.0622
s(Trial):Path-verb_Path-priming	3.3898	4.1975	33.4605	< 0.0001
s(Trial):Manner-verb_Baseline	1.0000	1.0001	1.8964	0.1685
s(Trial):Manner-verb_Manner-priming	1.5883	1.9542	9.1664	0.0163
s(Trial,Subject)	108.6682	782.0000	536.1778	< 0.0001
s(Item)	0.0001	31.0000	0.0001	0.7536

(Question 2)' in main document) was identical to the one for Question 1 (see Appendix S4). The only difference was that this model was fit to the L2 learner data. Function call in R:

```
mgcv::bam(Used ~ VbType_Cond + s(Trial, by = VbType_Cond) + s(Trial,
Subject, bs = 'fs') + s(VideoName, bs = 're'), data = d_l2, family =
'binomial')
```

Table S5 shows the summary of the GAMM for trial-by-trial adaptation in L2 learners. For a glossed example of the output, see Appendix S4.

Follow-up analysis

As noted in the main document, Figure 5 shows that L2 learners adapted to both path and manner priming over the course of the experiment. However, as was the case for Question 1 (see above), the GAMM does not allow us to numerically compare the strength of adaptation in each condition. If learners adapted more strongly to manner verbs, they would be behaving similar to native speakers, which in turn would suggest that their expectations are similar to those of native speakers. If, on the other hand, they either adapted more strongly to path verbs than to manner verbs, or equally strongly to both lexicalization types, it would suggest that learners at least partly transfer their L1-based expectations to their L2.

We proceeded analogously as for the follow-up analysis to Question 1, again with the caveat that this is a suboptimal approach (for the reasons explained above): To compare whether there was a difference in the strength of adaptation to path or manner verbs among learners, we selected the initial portion of the data that was approximately linear in log-odds in both priming conditions (i.e., trials 1 through 13, see Figure 5 in main document), and fitted a logistic GLMM. This model allows for an explicit statistical comparison of the slopes of interest (i.e., the slopes of the linear part of the curves in the lower panels of Figure 5). The model included as fixed effect predictors: Priming Condition (sum-coded: 1 = primed, −1 = baseline), Verb Type (sum-coded: 1 = manner verb, −1 = path verb), and Trial (trials 1 through 13, centred), as well as all their interactions. The random effect structure consisted of random by-subject and by-item intercepts, and a random by-subject slope for the centred Trial predictor, which was the random effect structure that was conceptually analogous to that of the GAMM. The function call and model specification in R was:

```
lme4::glmer(Used ~ Condition * VerbType * cTrial + (1 + cTrial |
Subject) + (1 | VideoName), data = d_l2_linear, family = 'binomial',
control = glmerControl(optimizer = 'bobyqa', optCtrl =
list(maxfun=2e5)))
```

The model output is shown in Figure S5. The main effect of Priming Condition was significant, indicating that primed participants used the verb they were primed with (path or manner verbs) more often than baseline participants (estimate = 1.43, $z = 4.97$, $p < .001$). The significant two-way interaction of Priming Condition and Trial further shows that primed participants increased their use of those verbs as the experiment proceeded more than did baseline participants (estimate = 0.19, $z = 4.05$, $p < .001$), which replicates what we saw from the GAMM (Figure 5 in main text). The critical fixed-effect coefficient in the model to assess if the strength of adaptation differed between priming conditions is the three-way interaction between Priming Condition, Verb Type and Trial. This coefficient indicates if the initial slope in the lower left panel of Figure 5 is steeper than the corresponding slope in the lower right panel of that figure (see main document). The coefficient was negative, indicating that numerically the slope was steeper for path-verb adaptation than for manner-verb adaptation, but it was non-significant (estimate = −0.05, $z = -1.09$, $p = .27$).

Thus, the way L2 learners adapted to the input is consistent with them basing their L2 expectations on a mixture of their L1 and L2 experience. Learners adapted to both lexicalization patterns as the experiment proceeded, but only showed a numerically (but not statistically significant) stronger adaptation to path verbs than to manner verbs.

```
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: binomial  ( logit )
Formula: Used ~ Condition_bin * VerbType * cTrial + (1 + cTrial | Subject) +      (1 | VideoName)
   Data: d_l2_lin
Control: glmerControl(optimizer = "bobyqa", optCtrl = list(maxfun = 2e+05))

     AIC      BIC   logLik deviance df.resid
   920.8    979.9   -448.4    896.8     1009

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.8260 -0.4225  0.0173  0.3795  3.3357

Random effects:
 Groups    Name        Variance Std.Dev. Corr
 Subject   (Intercept) 4.398302 2.09721
           cTrial      0.064757 0.25447  0.66
 VideoName (Intercept) 0.001306 0.03614
Number of obs: 1021, groups:  Subject, 79; VideoName, 32

Fixed effects:
                                                      Estimate Std. Error z value Pr(>|z|)
(Intercept)                                            0.26088    0.27722   0.941  0.34668
Condition_binPrimed_vs_Baseline                        1.43025    0.28790   4.968 6.77e-07 ***
VerbTypeM_vs_P                                        -1.55804    0.29289  -5.320 1.04e-07 ***
cTrial                                                 0.14573    0.04431   3.289  0.00100 **
Condition_binPrimed_vs_Baseline:VerbTypeM_vs_P        -0.17858    0.27660  -0.646  0.51852
Condition_binPrimed_vs_Baseline:cTrial                 0.19117    0.04715   4.054 5.03e-05 ***
VerbTypeM_vs_P:cTrial                                 -0.14354    0.04849  -2.960  0.00307 **
Condition_binPrimed_vs_Baseline:VerbTypeM_vs_P:cTrial -0.04813    0.04404  -1.093  0.27442
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
               (Intr) Cn_P__B VrTM__P cTrial Cn_P__B:VTM__P C_P__B:T VTM__P:
Cndtn_bP__B     0.090
VrbTypM_v_P    -0.104 -0.220
cTrial          0.578  0.138  -0.143
Cn_P__B:VTM__P -0.142 -0.096   0.082  -0.142
Cndt_P__B:T     0.130  0.613  -0.232   0.171 -0.115
VrbTyM__P:T    -0.132 -0.227   0.618  -0.161  0.114         -0.333
C_P__B:VTM__P: -0.141 -0.118   0.117  -0.219  0.581         -0.145    0.148
```

Figure S5. Follow-up analysis to Question 2. Complete output for the generalized linear mixed model fitted to the L2 speaker data (as obtained from the *summary* function in R).

Appendix S6. GAMM investigating the effect of L2 proficiency on adaptation (Question 3)

For the last analysis reported in the paper (see Results > Question 3, in main document), we added a new continuous predictor, L2 Proficiency Score (i.e., ClozeScore), to the GAMM fitted in Question 2 (see Appendix S5). This model again was fitted to the learner data only. L2 Proficiency Score was allowed to interact with Verb Type, Priming Condition, and Trial. We used a tensor product approach to smoothing, because this is the preferred option when the continuous variables are not measured on the same units (Wood, 2006), as is the case here. The function call and model specification in R for this GAMM were:

```
mgcv::bam(Used ~ VbType_Cond + te(Trial, ClozeScore, by = VbType_Cond)
+ s(Trial, Subject, bs = 'fs') + s(VideoName, bs = 're'), data = d_l2,
family = 'binomial')
```

The model output is shown in Table S6. For a glossed example of the output, we refer the reader to Appendix S4. Figures S6-1 and S6-2 show visualizations of the model: Figure S6-1 shows estimates for the use of path verbs in the path-primed and the baseline conditions, and Figure S6-2 shows estimates for the use of manner verbs in the manner-primed and baseline conditions. As we detail next, these figures corroborate the result reported in the main document (note that Figure 6 in the main document shows the *difference* between the two curves in each series of panels of Figures S6-1 and S6-2).

Table S6. Summary of the GAMM to assess the effect of L2 proficiency on trial-by-trial adaptation (Question 3). The data comes from L2 learners only. The explained deviance of the model is 55.2%.

A. parametric coefficients	**Estimate**	**Std. Error**	**z-value**	**p-value**
Intercept (Path-verb_Baseline)	0.5754	0.4004	1.4371	0.1507
Path-verb_Path-priming	2.4912	0.6165	4.0411	0.0001
Manner-verb_Baseline	-2.9459	0.5782	-5.0947	< 0.0001
Manner-verb_Manner-priming	-0.1991	0.5616	-0.3545	0.7230
B. smooth terms	**edf**	**Ref.df**	**F-value**	**p-value**
te(Trial,ProficiencyScore): Path-verb_Baseline	6.1700	7.7520	15.7228	0.0338
te(Trial,ProficiencyScore): Path-verb_Path-priming	8.4326	10.5423	36.0505	0.0001
te(Trial,ProficiencyScore): Manner-verb_Baseline	3.0000	3.0001	3.7955	0.2844
te(Trial,ProficiencyScore): Manner-verb_ Manner-priming	5.8462	7.3517	11.1586	0.1522
s(Trial,Subject)	96.7831	774.0000	469.8387	< 0.0001
s(Item)	0.0001	31.0000	0.0001	0.7612

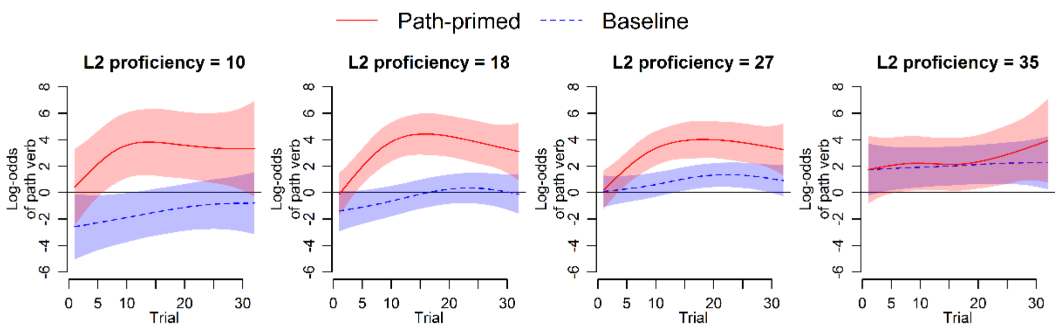

Figure S6-1. Visualization of path adaptation among learners as a function of L2 proficiency. The GAMM predicted the combined effects of Trial, L2 Proficiency Score and Condition (primed vs. baseline) on the log-odds of using a path or manner verb (as coded by the indicator variable Verb Type, which here is set to 'path verbs'). The panels show model estimates at different levels of L2 proficiency.

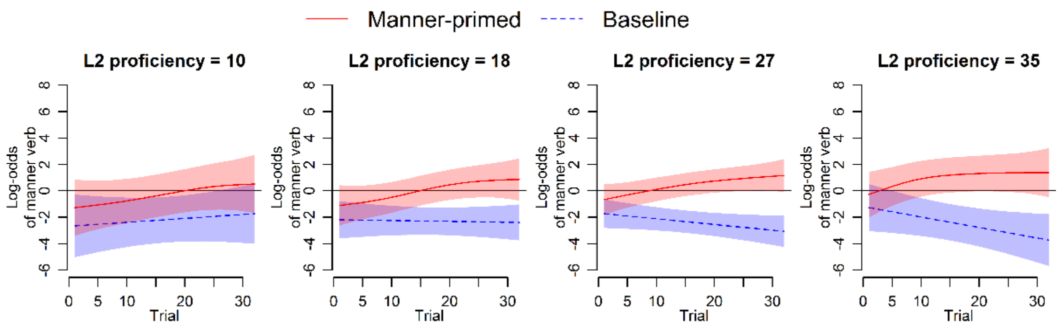

Figure S6-2. Visualization of manner adaptation among learners as a function of L2 proficiency. The GAMM predicted the combined effects of Trial, L2 Proficiency Score and Condition (primed vs. baseline) on the log-odds of using a path or manner verb (as coded by the indicator variable Verb Type, which here is set to 'manner verbs'). The panels show model estimates at different levels of L2 proficiency.

In what follows, we offer a detailed interpretation of the Figures S6-1 and S6-2. Starting with Figure S6-1, path adaptation (i.e., the *difference* between the two curves) was significant already at low proficiency scores (= 10). This is shown by the fact that the confidence bands, although broad (reflecting the reduced number of participants at this level of proficiency in our sample), do not overlap (a sufficient but not necessary condition to determine significance). At somewhat higher proficiency levels (=18), confidence bands become narrower (there is more data for these proficiency levels) and the quick adaptation effect becomes clearly visible, with the two lines diverging during the first trials. At proficiency scores of 27, however, the estimated adaptation effect (the difference between the curves) already becomes smaller. The adaptation effect for path finally disappears for the more proficient L2 speakers in our sample (proficiency scores of 35). The last plot looks qualitatively similar to that of native speakers (see Figure 4 in the main document, upper left panel). Note that the decreasing adaptation effect at higher proficiency levels (i.e., decreasing difference between the two curves) is due both to a flattening of the curve for path-primed participants (continuous red line) and a steady upwards shifting of the curve for participants in the baseline condition (dashed blue line).

We now turn to Figure S6-2, showing the model estimates for manner adaptation. It suggests that the effect of proficiency on adaptation is the opposite for manner priming than for path priming. The effect of Trial on the production of manner verbs in the baseline condition (blue dashed lines in the different panels) changed from a mildly positive slope to a mildly negative slope as a function of increasing L2 proficiency (although confidence bands remain broad at all levels). At the same time, the use of manner verbs in the manner-primed condition (red continuous line in the different panels) became more likely with growing L2 proficiency (see in particular the changes in the initial slope of the red continuous line across panels). Together, this led to increasing adaptation to manner verbs with growing L2 proficiency. That is, the *difference* between the dashed blue line and the continuous red line became larger from left to right).

As for path-verb adaptation, manner-verb adaptation suggests that L2 speakers' patterns of adaptation increasingly came to resemble those of native speakers with growing proficiency (see Figure 4 in the main document). Strikingly, this is the case even though the direction of the effects of proficiency on adaptation were the opposite for path and manner verbs.

Appendix S7. Self-priming

A reviewer brought to our attention the issue of self-priming, in the following way:

> Strictly speaking, each priming trial consists of two different types of utterances: the written and orally repeated sentence, and the sentence produced by the participant to describe the picture. I take it that in the adaptation analysis, only the number of trials has been taken into consideration (that is, each prime+target pair is counted as 1). Would the result be any different if in addition to the number of primes, also the number of path or manner constructions produced as descriptions by the participant were taken into consideration as well?

To gain at least some insight into the role of self-priming, we took the following approach:

1. We computed a new variable *SelfProduction* that counts, for each trial, the (cumulative) number of times a speaker has produced the target verb type (path or manner verb, depending on the value of the *VerbType* variable) in all trials up to the last one. This variable is always zero on the first trial and goes up to maximally 31 (in trial 32, the last trial) if the participant has produced the target verb throughout the whole experiment.
2. We fit a GAMM identical to the ones reported in the main paper, except we use *SelfProduction* as a predictor instead of *Trial*.

If self-priming can account for the results we observed, then we should see *no differences* between the prime and baseline groups. This is because, when using *SelfProduction* as the predictor, speakers in the different conditions are being treated the same: a speaker for whom, say, *SelfProduction* = 7 in a given trial, has been self-primed seven times, irrespective of whether he or she is in the baseline or in the primed condition. So if self-priming is all there is to the observed effects, then the two groups should be indistinguishable when the data is thus analyzed. Conversely, to the extent that we still see a difference between the conditions, this difference cannot be due to self-priming, but has to be attributed to the only factor that differed between conditions, namely the fact that primed participants also had to read and repeat the priming sentences.

Figures S7-1 through S7-4 below present the results of the re-analyses for Questions 1-3. The R code that generated the analyses and figures, as well as model summaries, is available as a Dataverse repository at https://doi.org/10.7910/DVN/TOJ1UH.

For all three questions, we found the same qualitative effects between conditions (primed vs baseline) when we control for participants' own production rather than the sentences they were primed with. Self-priming is thus insufficient to explain our results. Note that confidence intervals for high values of *SelfProduction* are wider than those of the analysis in the main text (which use trials as predictor). This is due to the fact that fewer subjects contribute data as the value of *SelfProduction* increases (e.g., only those participants who have described *all* pictures up to the last one with the primed structure contribute to 31 *SelfProductions*).

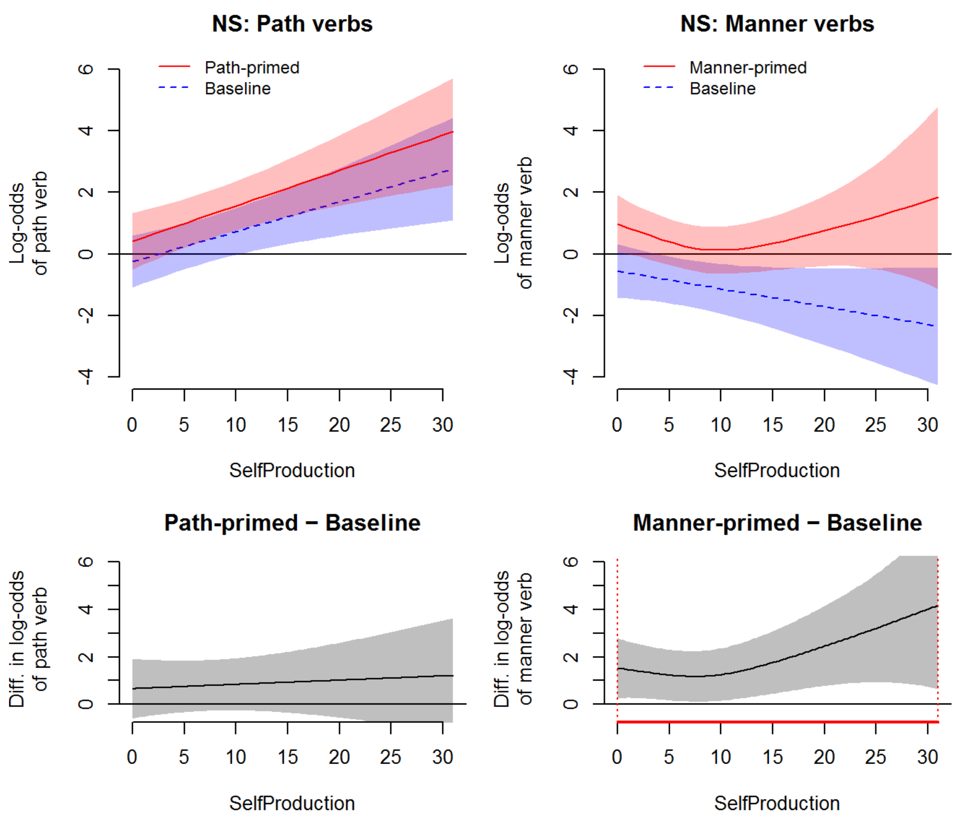

Figure S7-1. Self-priming: Reanalysis of Q1 replacing the variable Trial with the variable SelfProduction, which for each trial counts the number of times a participant produced the target structure up until the previous trial. See text for details.

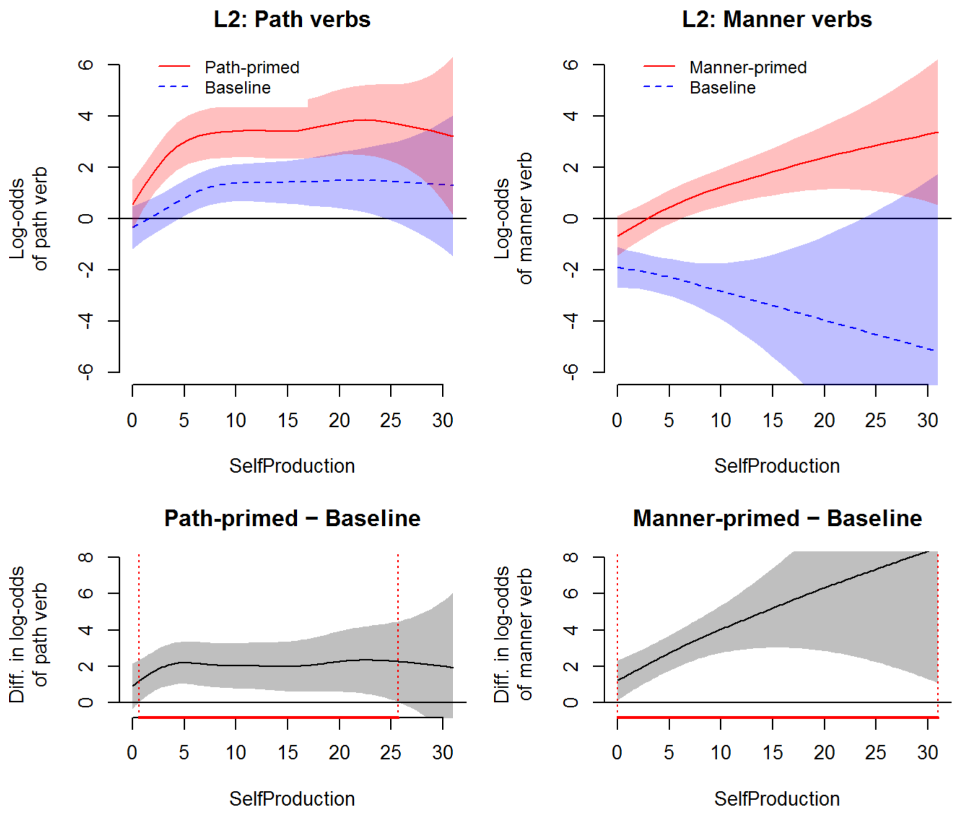

Figure S7-2. Self-priming: Reanalysis of data for Q2 replacing the variable Trial with the variable SelfProduction, which for each trial counts the number of times a participant produced the target structure up until the previous trial. See text for details.

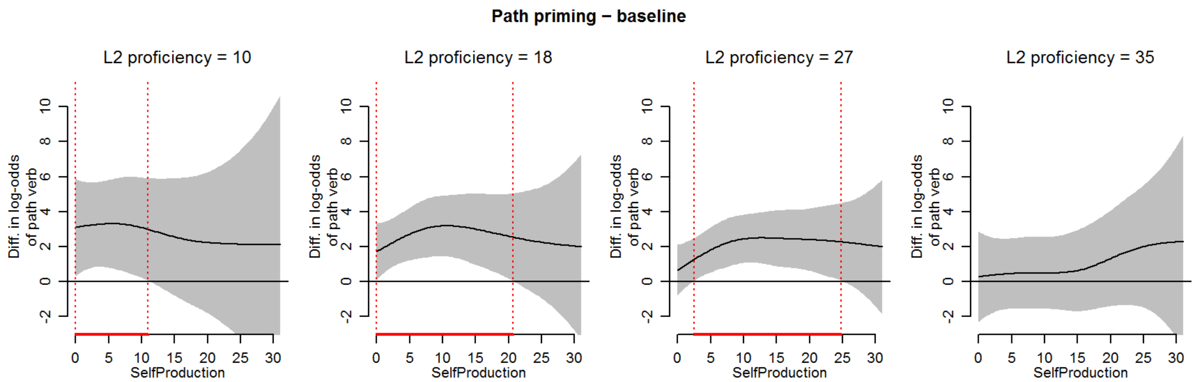

Figure S7-3. Self-priming: Reanalysis of data for Q3 (Path verb production), replacing the variable Trial with the variable SelfProduction, which for each trial counts the number of times a participant produced the target structure up until the previous trial. See text for details.

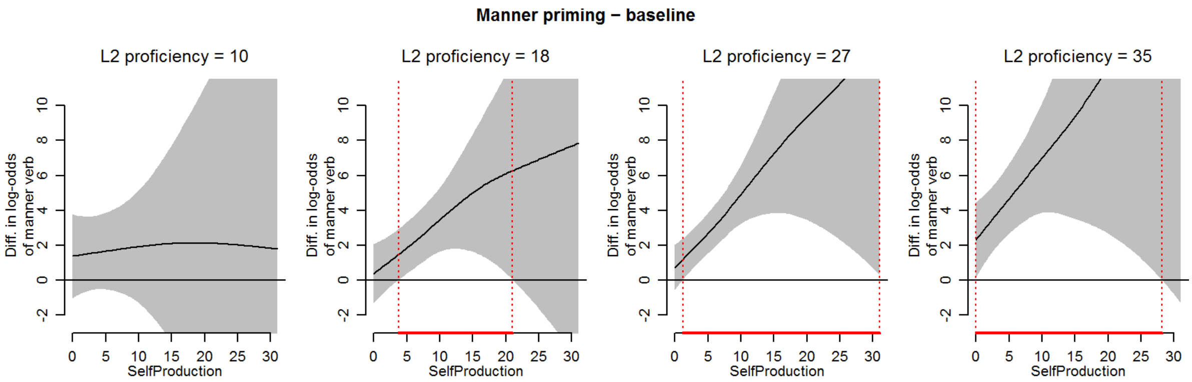

Figure S7-4. Self-priming: Reanalysis of data for Q3 (Manner verb production), replacing the variable Trial with the variable SelfProduction, which for each trial counts the number of times a participant produced the target structure up until the previous trial. See text for details.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Tremblay, A. (2011). Proficiency assessment standards in second language acquisition research: “Clozing” the gap. *Studies in Second Language Acquisition*, *33*(3), 339–372. https://doi-org.ezp.sub.su.se/10.1017/S0272263111000015

van Rij, J. (2015, March). Overview GAMM analysis of time series data. Retrieved 7 November 2018, from http://jacolienvanrij.com/Tutorials/GAMM.html

Vanhove, J. (2015). Analyzing randomized controlled interventions: Three notes for applied linguists. *Studies in Second Language Learning and Teaching*, *5*(1), 135–152.

Winter, B., & Wieling, M. (2016). How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling. *Journal of Language Evolution*, *1*(1), 7–18. https://doi.org/10.1093/jole/lzv003

Wood, S. N. (2006). *Generalized additive models: an introduction with R*. Boca Raton, FL: Chapman & Hall/CRC.

Wood, S. N. (2013). On p-values for smooth components of an extended generalized additive model. *Biometrika*, *100*(1), 221–228. https://doi.org/10.1093/biomet/ass048