When one probabilistic grammar is not restrictive enough: The curious case of an L2 learning simulation*

Silke Hamann
ACLC, University of Amsterdam

Chao Zhou
Centro de Linguística, Faculdade de Letras, Universidade de Lisboa

This article explores an intriguing case described by Zhou and Hamann (2026), who provided a Noisy Harmonic Grammar account for how perception and orthography together shape the L2 acquisition of the Spanish palatal approximant. Although the proposed grammar successfully generates the attested outputs in a categorical manner, an accompanying computer simulation showed a mismatch: a single grammar produced more variation than was observed empirically, while a two-grammar simulation exactly mirrored the data. Through a close analysis of the constraints and candidate evaluation, we lay bare the underlying contradictory requirements on the constraint weights, and demonstrate how such a contradiction can be identified and resolved by two separate grammars.

1 Introduction

Constraint-based phonological theories such as Optimality Theory (OT; Prince & Smolensky 1993/2004) and Harmonic Grammar (HG; Legendre, Miyata & Smolensky 1990) are appealing to first and second language researchers as they allow the computational simulation of language acquisition through associated learning algorithms. Such learning algorithms, equipped with a constraint set and a set of language data (usually in the form of input-output pairs, resembling the input-output distribution frequency attested empirically), aim to find a ranking or weighting of constraints (that is, a grammar) that can account for the given dataset.

While early learning algorithms such as Tesar's (1995) Constraint Demotion still operated on the idea that languages show no variation, the advent

* This article emerged from a discussion with Paul on the results of one of the simulations we performed in Zhou and Hamann (2026), which we found problematic, while Paul was fascinated with this side finding. As always, we are deeply indebted to him for his feedback and discussions.

© Silke Hamann & Chao Zhou

of stochastic OT (Boersma 1997) and the Gradual Learning Algorithm (GLA; Boersma 1997, Boersma & Hayes 2001) changed this view: With its probabilistic decision mechanism and the added noise at each evaluation time, it allowed researchers to account for non-categorical behaviour in a single grammar. More recently, a stochastic version of HG, noisy HG, with its associated learning algorithm (Boersma & Pater 2016), has been used in learning simulations, successfully yielding outcomes that mirror the variation observed in the empirical data. Most studies adopting stochastic OT or noisy HG have attempted to account for probabilistic output distributions within a single grammar, arguably reflecting a cautious preference for a more restrictive approach which does not assume that speakers of a language may have different phonological grammars. Zhou and Hamann (2026), however, present a case that challenges this single-grammar approach. Revisiting an L2 acquisition experiment (data by Rafat & Stevenson 2019), they demonstrate that the observed variation in the learners' output cannot be modelled with a single noisy HG grammar, as this resulted in an unattested winning form. Instead, the experimental findings could only be replicated by combining the results of two grammars, representing two learner types. In the present article, we further investigate this finding to better understand the underlying mechanisms that require the postulation of two (or more) learner types in a learning simulation with a stochastic learning algorithm.

The article is structured as follows. We briefly introduce the case study in section 2. In section 3 we isolate the variation that could not be captured in one grammar, and in section 4 we show how this variation can be assigned to two learner types. Section 5 concludes.

2 A curious simulation case

The data for our learning simulation stem from an experiment by Rafat and Stevenson (2019), where speakers with Canadian English as their first language (L1) who had no previous knowledge of Spanish were auditorily exposed to the Mexican Spanish palatal approximant /j/ (which is often realized as fricative or affricate, Avelino 2018). The sound occurred word-initially or -medially, as in [janeɾo] and [majeto], and was presented first without and then with orthographic representations, as in <llanero> and <malleto>. In the auditory-only condition, the participants perceived the [j] sound always as /j/. When simultaneously exposed to the orthography, participants still perceived mostly a /j/, but also /l/, indicating that they paid attention only to the orthography, and /lj/, exhibiting an integration of both auditory and orthographic cues.

In Zhou and Hamann (2026), we formalized this and another case of an intriguing interaction between perception and orthography in the HG version of *Bidirectional Phonology and Phonetics* (BiPhon, Boersma 2007; see e.g.

Hamann, Apoussidou & Boersma 2012 for a BiPhon-HG implementation). In addition to auditory and phonological surface forms, the formalization employed orthographic representations and constraints that map orthographic onto surface phonological forms (Hamann & Colombo 2017). To achieve greater cognitive plausibility and to rely on fewer innate assumptions, all constraints were formulated positively (Boersma & Pater 2007) and the evaluation rewarded constraint satisfaction instead of punishing constraint violation (see Zhou & Hamann 2026, for a detailed discussion).

To test the validity and learnability of the proposed constraints and constraint weights, we furthermore provided learning simulations in Praat (Boersma & Weenink 2026), employing OTGrammar and PairDistribution objects. The grammar script was set to a Noisy-HarmonicGrammar decision strategy. Initial constraint weights were assigned a numerical value of 20 rather than the standard weight of 100, as the latter introduces a larger bias to maximize the number of satisfactions (Boersma, p.c.), and a first simulation with starting weights of 10 resulted in a mixture of positive and negative constraint weights, which we deemed undesirable. All other parameters in the learning simulations were set to the default values in Praat. Note that in the following tableaux, constraint satisfaction is indicated by "+1", even though in the Praat scripts this has to be indicated with a "−1" (as the default "1" is reserved for a constraint violation).

The learning distribution included one input form per condition (auditory only or auditory plus orthography) and position (word-initially or -medially), resulting in four input forms, see the first column of Table 1. The output forms and occurrence frequencies, given in the second and third column of Table 1, were based on the experimental findings by Rafat and Stevenson (2019), though this source only provides overall percentages that are not split by position. We assumed for the auditory plus orthography condition that the form /l/ occurs as alternative to /j/ only in initial position, as in medial position the candidate /ma.le.to/ is harmonically bounded by /mal.je.to/, hence can never win.

This first simulation resulted in a grammar that did not fully reflect the observed results, as the winner for the auditory plus orthography condition in medial position was in 4% of the cases the candidate /maj.le.to/, a form of which Rafat and Stevenson explicitly say that it never occurred in their experimental findings (2019: 445).

This discrepancy between observed data and simulation results could stem from the use of an inappropriate constraint set. However, all constraints in our simulation were independently motivated as each played an irreplaceable role in the learners' L1 English perception and reading grammar. At the same time, a formalization of the observed phenomena with categorical results (not taking account of the variation) showed that this set of constraints was sufficient. A

change in constraints (by e.g. incorporating additional constraints or redefining existing ones) therefore is not expected to solve the problem.[1]

The fact that the experimental results of the other case study (Zhou & Hamann 2020) that we modelled in Zhou and Hamann (2026) showed not only "cue integrators", but also learners that totally ignored the perceptual cues, or those that totally ignored the orthographic information, led us to assume that Rafat and Stevenson's experimental results probably also reflected more than one learner type. We could determine that the results can be accounted for if they are stemming from the combination of a learner type I, who does not pay any attention to orthography, and a learner type II, who shows integration of auditory cues and orthographic information in medial position, and who varies between cue and orthography only in initial position (where the cue integration is prevented due to a structural restriction that /lj/ is not a valid onset in Canadian English, while in medial position this illegal sequence could be made legal by an intervening syllable boundary: /l.j/). The input-output distribution frequencies for these two learner types are provided in the last two columns of Table 1. To account for the observed results, one needs to assume that the population of learners consists of 86% type I and 14% type II learners.

Table 1: The input-output distributions for the simulations. The third column is the total occurrence frequencies based on Rafat & Stevenson (2019), which was used in the first simulation. Columns four and five give the frequencies used in the second simulation for the two learner types I and II (and how often the learner types occurred), who together result in the same total frequency.

Input	**Output**	**Occurrence frequency**	**Type I (86%)**	**Type II (14%)**
[janeɾo]	/ja.ne.ɾo/	100%	100%	100%
[janeɾo] <llanero>	/ja.ne.ɾo/	93%	100%	50%
	/la.ne.ɾo/	7%	0%	50%
	/lja.ne.ɾo/	0%	0%	0%
[majeto]	/ma.je.to/	100%	100%	100%
[majeto] <malleto>	/ma.je.to/	86%	100%	0%
	/ma.le.to/	0%	0%	0%
	/mal.je.to/	14%	0%	100%

[1] A potential alternative explanation for the discrepancy between observed data and our first simulation could be that all constraints had the same initial weights, which is unrealistic, as L2 learners normally start with their L1 grammar, i.e., with constraint weights acquired on the basis of L1 input, and adjust it as a function of the L2 input they receive. A simulation with different initial constraint weights would, however, not resolve the problem. Section 3 proves that the attested variation can never be captured by a single grammar, regardless of its initial state.

A second simulation with two grammars and the two input-output distributions for type I and type II resulted in two stable grammars that combined in the ratio mentioned above and could exactly reproduce the observed output frequencies. No undocumented winning candidates emerged in either of these grammars.

3 Variation that one simulated learner cannot reproduce

After having laid out the failure of the first simulation and the success of the second, we will now try to lay bare the contradiction that resulted in a non-attested form in the first simulation. Tableaux 1a–1d provide all four input forms with their candidates, constraint satisfactions, and expected output distributions. There are four constraints: a structural constraint C_1 (avoiding onset /lj/ and /jl/, summarised as "ENGLISH PHONO(TACTICS)"), two cue constraints C_2 and C_3, of which C_2 is mapping the glide-vowel transitions present in all input forms onto a sequence of a glide and a vowel in the surface form, and C_3, mapping the vowel-glide transitions only present in intervocalic position (Tableaux 1c and 1d) onto a surface vowel-glide sequence, and finally an orthographic constraint C_4, mapping written <ll> onto surface /l/, which is only relevant in the inputs that also contain an orthographic form (Tableaux 1b and 1d). For full definitions of the constraints, see Zhou and Hamann (2026).

Tableau 1a: Auditory-only categorisation of word-initial glide

[janero]	C_1 ENGLISH PHONO	C_2 [jV] /jV/	C_3 [Vj] /Vj/	C_4 <ll> /l/
100% ☞ /ja.ne.ɾo/	+1	+1		
/la.ne.ɾo/	+1			
/lja.ne.ɾo/		+1		
/jla.ne.ɾo/				

Tableau 1b: Auditory-plus-orthography categorisation of word-initial glide

[janero] <llanero>	C_1 ENGLISH PHONO	C_2 [jV] /jV/	C_3 [Vj] /Vj/	C_4 <ll> /l/
93% ☞ /ja.ne.ɾo/	+1	+1		
7% ☞ /la.ne.ɾo/	+1			+1
/lja.ne.ɾo/		+1		+1
/jla.ne.ɾo/				+1

Tableau 1c: Auditory-only categorisation of word-medial glide

[majeto]	C_1 ENGLISH PHONO	C_2 [jV] /jV/	C_3 [Vj] /Vj/	C_4 <ll> /l/
100% ☞ /ma.je.to/	+1	+1	+1	
/ma.le.to/	+1			
/ma.lje.to/		+1		
/ma.jle.to/			+1	
/maj.le.to/	+1		+1	
/mal.je.to/	+1	+1		

Tableau 1d: Auditory-plus-orthography categorisation of word-medial glide

[majeto] <malleto>	C_1 ENGLISH PHONO	C_2 [jV] /jV/	C_3 [Vj] /Vj/	C_4 <ll> /l/
86% ☞ /ma.je.to/	+1	+1	+1	
/ma.le.to/	+1			+1
/ma.lje.to/		+1		+1
/ma.jle.to/			+1	+1
/maj.le.to/	+1		+1	+1
14% ☞ /mal.je.to/	+1	+1		+1

A closer look at the constraint satisfactions in each case provides information on the weights these constraints have in relation to each other. In Tableau 1a, there is only one winner, satisfying both C_1 and C_2; all other candidates are harmonically bounded by the winner because they have fewer satisfactions. No conclusions on the relative weights can be drawn from this. Similarly inconclusive is the evaluation of the form in Tableau 1c, where the first, winning candidate harmonically bounds all other candidates.

From Tableau 1b we can conclude that C_1 must have a higher weight than C_2, because the second candidate wins in 7% of the cases, and is distinguished from the never-winning third candidate in satisfying C_1 and not C_2. And C_4 must have a weight that is slightly lower but close enough to C_2 to result in a winning first candidate (satisfying C_1 and C_2) in 93% and a winning second candidate (satisfying C_1 and C_4) in 7% of the cases due to stochastic evaluation noise. These weight relations are captured in (1a).

Another relative weight relation can be found in the evaluation in Tableau 1d. Here, the sixth candidate wins in 14% of the cases, and differs from the fifth candidate, which never wins, by satisfying C_2 but not C_3. This indicates that C_2 must have a considerably higher weight than C_3. And a comparison between the first, often winning candidate and the sixth, less often winning, suggests that the

weight of C_3 can only be slightly higher than that of C_4. These two observations are captured in (1b).

(1) Overview of weight relations: " >>" stands for higher weight, "≥" slightly higher but close enough to allow for a reverse order in stochastic evaluation

a) $C_1 >> C_2 \geq C_4$
b) $C_2 >> C_3 \geq C_4$

A comparison between the two observed weight relations in (1) reveals the contradiction: While (1a) requires C_2 and C_4 to be fairly close in weight to result in C_4 occasionally having a higher weight due to evaluation noise, (1b) requires C_2 to have a far higher weight than C_4 (with the weight of another constraint, C_3, in between). These two contradicting requirements cannot be fulfilled by one grammar.

A solution to this problem is to assign these contradictory requirements to two different grammars. This is also what we devised when deciding to have two learner types that treat orthographic information differently. Learner type I ignores orthography by having lower weight for the orthographic constraint C_4 than for the cue constraint that occurs in all contexts C_2, thereby incorporating requirement (1b). Learner type II, on the other hand, gives orthography and its respective constraint C_4, similar weight as the context-independent cue constraint C_2, and satisfies requirement (1a). The actual weights from the computer simulation with these two learner types from Zhou and Hamann (2026) are given in Table 2.

Table 2: Weights of the four constraints for the grammars of the two learner types (first and second row) in the second simulation, and for the one, compromise grammar in the first simulation (third row).

		C_1	**C_2**	**C_3**	**C_4**
Simulation 2	Type I	23	24	23	10
	Type II	30	21	8	21
Simulation 1		22	21	20	17

Returning to the contradictory requirements in constraint weights for the two input evaluations in Tableaux 1b and 1d, these can be simplified as in Tableau 2: two input evaluations, with two constraints, and two output candidates each satisfying the same constraints.

Tableau 2: Simplification of variation and constraint satisfactions of Tableaux 1b and 1d

input 1	C_a	C_b
93% ☞ output 1a	+1	
7% ☞ output 1b		+1
input 2		
100% ☞ output 2a	+1	
output 2b		+1

While variation in the output for input 1 in Tableau 2 requires the two constraints to have almost similar weights, no variation in the output to input 2 requires them to have weights that are far enough apart to never result in a change due to stochastic evaluation. Note that the output variation that we observed in Tableau 1d between the first and sixth candidate was due to an additional constraint (C_3) that does not seem to play a role in the weight contradiction and therefore is neglected for now.

A computer simulation with input forms, constraints, constraint satisfactions and input-output distributions based on Tableau 2 attested that this reduction of the problem is correct, because it replicated the erroneous pattern of a winning output 2b: Both output forms 1b and 2b occurred 3.5% of the time and both output forms 1a and 2a 96.5% of the time. Since the two output pairs exhibit exactly the same constraint satisfactions, one probabilistic grammar can only result in the same output distributions for both. Learning pairs of input 1 – output 1b, which occur 3.5% of the time of all learning pairs, cause the GLA to add weight to constraint C_b. Learning pairs of input 2 – output 2b, on the other hand, have a distribution of 0% in the Pair distribution because they never occur in the data, and are therefore not fed to the GLA. Their non-occurrence cannot cause a lowering of the weight of constraint C_b, as would be required to counteract the learning based on input 1 – output 1b pairs. As a result, the variation learned for input 1 is automatically transferred to input 2.

There is no parameter setting in a GLA learning simulation that allows for a different ranking of two constraints between two input forms within one grammar. One might wonder whether changing the value for the evaluation noise that is added to constraint weights at each evaluation time would lead to a difference in output distribution between the two input forms in Tableau 2. We tested this possibility by running two further simulations with different noise values. Higher evaluation noise results in a larger gap between the constraint weights of C_a and C_b, and lower noise in a smaller gap, see the results in Table 3, but neither allows for a difference in output distribution between the two input

forms. Interestingly, a change in evaluation noise does result in output distributions that are different from the input distributions; presumably further settings (such as initial plasticity, replication per plasticity, plasticity decrement, number of plasticities and relative plasticity spreading) have to be adapted to attain an equilibrium comparable to that of the default settings, but this goes beyond the scope of the present paper.

Table 3: Average weights of the two constraints (first two rows) and output distributions for the four candidates (rows 3–6) after 100 learning simulations each with three different evaluation noise settings.

	Evaluation noise		
	2 (default)	**0.5**	**5.0**
Weight C_a	22.6	20.7	26.4
Weight C_b	17.4	19.4	13.6
Output 1a	96.5%	67.5%	99.9%
Output 1b	3.5%	32.5%	0.1%
Output 2a	96.5%	67.5%	99.9%
Output 2b	3.5%	32.5%	0.1%

The same constraint values and output distributions were obtained when the decision mechanism was set to Optimality Theory instead of Harmonic Grammar, i.e. the constraints had strict rankings rather than weights.

The presented simplification of the contradiction to two constraints can help detect other cases of contradictory constraint weight requirements in conglomerated data (before or after learning simulations): If two constraints are determining some kind of output variation for one input, and at the same time are responsible for categorical output for another input, this is an indication that different types of learners are conglomerated in one data set.

4 Different learner types

As shown in sections 2 and 3, the contradiction on the constraint weights (or rankings) in the case simulated in Zhou and Hamann (2026) could be solved by assuming that the overall output distribution does not stem from one learner or one homogeneous group of learners, but from two different types: The first shows categorical behaviour for both input 1 and 2 in only having output a as winner, see Type I in Table 4. The variation that can be found for input 1 is then taken care of by the less-frequent learner Type II, who varies between the two outputs, see last column in Table 4. However, this learner cannot show the same variation

for input 2, as this would result in the occurrence of output 2b, which is not observed in the data.

Table 4: The input-output distributions for two potential learner types for the simplified case.

Input	Output	Occurrence frequency	Type I (86%)	Type II (14%)
1	1a	93%	100%	50%
	1b	7%	0%	50%
2	2a	100%	100%	0%
	2b	0%	0%	0%

What our simplification predicts is that learner type II uses neither output 2a nor 2b for input form 2 (otherwise the same variation would be expected as for input 1 due to the weights of the relevant constraints). Does that mean that learner type II does not produce an output for input 2? This is not likely. Instead, a third output form 2c is expected to be produced by this learner, one satisfying a constraint that did not play a role in the evaluation of input 1. This is also what we observed in the simulation in section 2, where a constraint C_3, the cue constraint that only applies in medial position and that we ignored in the detection of the weight contradiction in section 3, plays a role: This constraint allowed for a different winning candidate in the evaluation in Tableau 1d, here input form 2, compared to the evaluation in Tableau 1b, here input form 1.

A summary and generalisation of these observations to other frequency ratios is given in Table 5:

Table 5: The input-output distributions for two potential learner types that are conglomerated in the observed data (total frequency), where *x* and *y* are variables for the occurrence frequencies of the two types of learners.

Input	Output	Constraint satisfied	Total frequency	Type I (*x*%)	Type II (*y*%)
1	1a	C_1	$(x+\frac{1}{2}y)$%	100%	50%
	1b	C_2	$(\frac{1}{2}y)$%	0%	50%
	1c	C_3	0%	0%	0%
2	2a	C_1	*x*%	100%	0%
	2b	C_2	0%	0%	0%
	2c	C_3	*y*%	0%	100%

In this table, we assume that for learner type II, output forms 1a and 1b occur similarly often, but this ratio could also vary.

5 Conclusion

In this paper, we explored an intriguing learnability issue detected by Zhou and Hamann (2026): not all group-level variation observed in empirical data can be computationally modelled by a single probabilistic HG or OT grammar. We first demonstrated how the logical contradiction implicated in the use of a single grammar to simulate two distinct input-output distributions can be identified. Given that the proposed constraints were able to account for categorical behaviour, we contended that simply adopting a different constraint set is unlikely to solve the problem. Additional learning simulations further revealed that the gap between single-grammar modelling and empirical evidence persists even when varying evaluation noises are introduced. Instead, the variation in question is best captured by positing (at least) two learner types, which may better reflect the empirical reality, given the well-documented prevalence of individual differences in speech learning, in L2 but also in L1 (see Kidd & Donnelly 2020 for a recent overview).

6 References

Avelino, H. (2018). Mexico City Spanish. *Journal of the International Phonetic Association 48*(2), 223–230. https://doi.org/10.1017/S0025100316000232

Boersma, P. (1997). How we learn variation, optionality, and probability. *Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam 21*, 43–58. https://www.fon.hum.uva.nl/archive/1997/1997-ifaproc21-Boersma_variation.pdf

Boersma, P. (2007). Some listener-oriented accounts of h-aspiré in French. *Lingua 117*(12), 1989–2054. https://doi.org/10.1016/j.lingua.2006.11.004

Boersma, P., & Hayes, B. (2001). Empirical tests of the Gradual Learning Algorithm. *Linguistic Inquiry 32*(1), 45–86. https://doi.org/10.1162/002438901554586

Boersma, P., & Pater, J. (2007). Constructing constraints from language data: the case of Canadian English diphthongs [Conference presentation]. *NELS 38*. https://www.fon.hum.uva.nl/paul/presentations/CanadianRaisingNELS2007.pdf

Boersma, P., & Pater, J. (2016). Convergence properties of a gradual learning algorithm for Harmonic Grammar. In J. McCarthy & J. Pater (eds.), *Harmonic Serialism and Harmonic Grammar,* 389–434. Equinox. http://dx.doi.org/10.3138/9781781793817-014

Boersma, P., & Weenink, D. (2026). *Praat: doing phonetics by computer* [Computer program] (Version 6.4.60). https://www.fon.hum.uva.nl/praat/

Hamann, S., Apoussidou, D., & Boersma, P. (2012). Modelling the formation of phonotactic restrictions across the mental lexicon. *Proceedings of the 45th Meeting of the Chicago Linguistic Society 45*(1), 193–206.

Hamann, S., & Colombo, I. E. (2017). A formal account of the interaction of orthography and perception. *Natural Language & Linguistic Theory 35*(3), 683–714. https://doi.org/10.1007/s11049-017-9362-3

Kidd, E., & Donnelly, S. (2020). Individual differences in first language acquisition. *Annual Review of Linguistics 6*(1), 319–340. http://dx.doi.org/10.1146/annurev-linguistics-011619-030326

Legendre, G., Miyata, Y. & Smolensky, P. (1990). Can connectionism contribute to syntax? Harmonic Grammar, with an application. In M. Ziolkowski, M. Noske & K. Deaton (Eds.), *Proceedings of the 26th Regional Meeting of the Chicago Linguistic Society*, 237–252. Chicago Linguistic Society.

Prince, A. & Smolensky, P. (2004). *Optimality Theory: constraint interaction in generative grammar.* Blackwell [Original work published 1993]. https://doi.org/10.1002/9780470756171.ch1

Rafat, Y., & Stevenson, R. A. (2019). Auditory-orthographic integration at the onset of L2 speech acquisition. *Language and Speech 62*(3), 427–451. https://doi.org/10.1177/0023830918777537

Tesar, B. (1995). *Computational Optimality Theory*. [PhD dissertation, University of Colorado]. (ROA 90).

Zhou, C., & Hamann, S. (2020). Cross-linguistic interaction between phonological categorization and orthography predicts prosodic effects in the acquisition of Portuguese liquids by L1-Mandarin learners. *Proceedings of Interspeech 2020*. Shanghai, 4486–4490. http://doi.org/10.21437/Interspeech.2020-2689

Zhou, C., & Hamann, S. (2026). A revised multi-modal formal account for orthography-induced sound patterns. *Linguistic Inquiry*. Advance online publication. https://doi.org/10.1162/LING.a.585

Silke Hamann, silke.hamann@uva.nl

Chao Zhou, zhouchao@edu.ulisboa.pt