International Conference on Error-Driven Learning in Language (EDLL 2021)

Wednesday, March 10 — Friday, March 12, 2021

University of Tübingen

Book of Abstracts

Edited by Jessie S. Nixon, Elnaz Shafaei-Bajestan and Harald Baayen

Foreword

This volume includes the proceedings abstracts of the International Conference on Error-Driven Learning in Language, EDLL 2021, held online, Tübingen, Germany, 10-12 March, 2021. This is the first conference of its kind.

We believe error-driven learning models have made and will likely continue to make an important contribution to our understanding of language. Therefore, it is with great pleasure that we bring together researchers working in this exciting area for the first conference dedicated to error-driven learning in language. We are happy to have such a great line up of Keynote speakers: Randall O'Reilly, Adele Goldberg and Petar Milin. We have papers on various levels of linguistic processing, from speech sounds to morphology, syntax and semantics, on first and second language acquisition, perception, comprehension and production, neural correlates of error, as well as investigations into specific details of the learning algorithms. We have participants joining us from around the globe. We would like to thank all the participants and presenters for their contributions. We couldn't have done it without you. Thanks also to Adnane Ez-Zizi for help with the programme and to the ERC for financial support (grant number 742545). Finally, we look forward to further research and discussion into the future!

Jessie Nixon
Elnaz Shafaei-Bajestan
Harald Baayen

Tübingen, March 2020

March 10, 2021

9:00 - 9:15	**COFFEE/TEA**	
9:15 - 9:30	**WELCOME**	
KEYNOTE 1		
9:30 - 10:30	Randall O'Reilly	Predictive Error-driven Learning in the Brain. y
10:30 - 10:40	**BREAK**	
10:40 - 11:10	Special event for junior researchers: Q&A with Randall O'Reilly	
11:10 - 11:15	**BREAK**	
SESSION 1		
11:15 - 11:45	**Yung Han Khoe, Chara Tsoukala, Gerrit Jan Kootstra and Stefan Frank**	Error-driven learning as a mechanism for cross-language structural priming y
11:45 - 12:15	**Chiara Gambi and Katherine Messenger**	The role of prediction errors in 4-year-olds' learning of English direct object datives y
12:15 - 12:45	**Jessica Nieder, Fabian Tomaschek and Ruben van de Vijver**	Modeling Maltese broken and sound plurals with Naive Discriminative Learning y
12:45 - 13:00	**BREAK**	
SESSION 2		
13:00 - 13:30	**Ben Ambridge and Kristen Liu**	Balancing information-structure and semantic constraints on construction choice: A discriminative learning model of passive and passive-like constructions in Mandarin Chinese (and Balinese and Hebrew). y
13:30 - 14:00	**Dušica Filipović Đurđević**	Uncertainty of polysemous word senses in the light of discrimination learning y
14:00 - 14:30	**Ksenija Mišić, Dušica Filipović Đurđević**	Interaction of semantic and syntactic ambiguity in the light of discrimination learning y
14:30 - 14:45	**BREAK**	
14:45 - 15:45	**POSTER SESSION**	

March 11, 2021

	SESSION 3	
14:30 - 15:00	**Ronny Bujok, Sybrine Bultena, James McQueen and Mirjam Broersma**	Accent Adaptation through Error-Based Learning y
15:00 - 15:30	**Kristin Lemhöfer**	Error-driven learning in L2 vocabulary and syntax: ERP correlates y
15:30 - 15:45	**BREAK**	
	SESSION 4	
15:45 - 16:15	**Sanne Poelstra, Jessie S. Nixon and Jacolien van Rij**	Does learning occur in the absence of cues? y
16:15 - 16:45	**Adnane Ez-Zizi, Dagmar Divjak and Petar Milin**	Error-correction mechanisms in language learning: tracking individual differences y
16:45 - 17:15	**Vsevolod Kapatsinski**	When backward transitional probabilities can be learned using forward prediction y
17:15 - 17:30	**BREAK**	
	KEYNOTE 2	
17:30 - 18:30	**Adele E. Goldberg**	Explain me this: Coverage encourages generalization and Statistical Preemption constrains it y
18:30 - 18:40	**BREAK**	
18:40 - 19:10	Special event for junior researchers: Q&A with Adele E. Goldberg	

March 12, 2021

SESSION 5		
9:30 - 10:00	**Theres Grüter, Yanxin (Alice) Zhu and Carrie N. Jackson**	Can forcing second language learners to generate prediction errors increase learning? y
10:00 - 10:30	**Yanxin (Alice) Zhu, Yang Zhao and Theres Grüter**	Second language learners' sensitivity to competing alternatives is modulated by proficiency: Evidence from L2 Mandarin y
10:30 - 11:00	**Chi Zhang and Min Wang**	Effects of input type frequency on structural priming and statistical preemption in the acquisition of L2 dative construction y
11:00 - 11:15	**BREAK**	
SESSION 6		
11:15 - 11:45	**Harish Tayyar Madabushi, Dagmar Divjak and Petar Milin**	Less is more? Language learning, between simple and deep embeddings y
11:45 - 12:15	**Laurence Romain, Petar Milin and Dagmar Divjak**	Learnability and Tense Aspect combinations in English: unveiling a dual system grounded in experience y
12:15 - 12:45	**Benjamin Tucker, Dagmar Divjak and Petar Milin**	A learning perspective on the emergence of abstractions y
12:45 - 13:00	**BREAK**	
KEYNOTE 3		
13:00 - 14:00	**Petar Milin**	What can be used from learning? y
14:00 - 14:15	**CLOSING REMARKS**	
14:15 - 14:20	**BREAK**	
14:20 - 14:50	Special event for junior researchers: Q&A with Petar Milin	
14:50 - ...	**SOCIAL EVENT**	

Poster session

Booth #1	**Jessie S. Nixon and Fabian Tomaschek**	Infant speech acquisition through error-driven learning of the acoustic speech signal. y
Booth #2	**Marion Coumel, Ema Ushioda and Kate Messenger**	Can error-based models account for language processing via syntactic priming? Investigating the effects of task and learner characteristics y
Booth #3	**Kun Sun**	Semantic Vectors Based on Discriminative Learning as Predictive of Lexical Psychological Properties y
Booth #4	**Maja Linke and Michael Ramscar**	How Distributional Context Solves the Variance Problem in Speech Sampling y
Booth #5	**Michaela M. Vann, Giulia Bencini and Virginia Valian**	Learning trajectories in L2 and bilingual language development: a structural priming investigation y
Booth #6	**Xuefeng Luo, Yu-Ying Chuang and Harald Baayen**	Linear Discriminative Learning in Julia y
Booth #7	**Motoki Saito, Fabian Tomaschek and R. Harald Baayen**	Triphone meanings co-determine tongue shape during articulation: An ultrasound study y

Predictive Error-driven Learning in the Brain

Randall O'Reilly
University of California, Davis; oreilly@ucdavis.edu

I will present some recent computational models of brain circuits that can support predictive error-driven learning, along with a discussion of prior work on how the brain might support something like error backpropagation more generally. Error backpropagation is the engine of modern deep neural network models, and there has been a bit of a resurgence of interest in its possible biological basis recently. Top-down connections in the cortex can potentially provide a mechanism of error propagation, and there are various proposals that make distinct biological predictions, which will be reviewed. Predictive learning provides an attractive solution to a remaining challenge: where do all the error signals come from in the first place? Specific circuits between the thalamus and cortex appear ideally configured to support a form of predictive learning, which differs significantly from other machine-learning / Bayesian approaches. Our models show that this mechanism can learn abstract categorical representations from movies of rotating and translating 3D objects, and captures classic statistical learning phenomena in speech recognition. It is also consistent with how many current deep neural network models are trained.

Explain me this:
Coverage encourages generalization and Statistical Preemption constrains it

Adele E. Goldberg
Princeton University; adele@princeton.edu

How is that native English speakers find novel patterns such as *She tweeted them the story* unremarkable but stubbornly judge *She explained him the story* unacceptable? This apparent paradox is addressed by recognizing speakers' goal: to express their intended messages while obeying the conventions of their language community. Experimental evidence indicates that productivity is encouraged by Coverage (roughly the extent to which the required generalization has been previously attested), while productivity is constrained by statistical preemption: the existence of a more conventional, accessible alternative.

What can be used from learning?

Petar Milin
p.milin@bham.ac.uk
University of Birmingham UK

In my talk I will present work done with the Out Of Our Minds team [outofourminds.bham.ac.uk]. Through selected case studies I aim to show the range and reach that learning has in our research. Firstly, as a starting point, we see the role of learning along the lines of Poggio's revision of Marr's levels of understanding *any* complex system, language included: "understanding at the level of learning is […] perfectly adequate as an explanation all by itself" (2012, p. 1019). But the question about what learning is, more specifically, still needs an answer.

Many would accept that *learning is* (relatively permanent) *change*. Such a broad proposal allows diverse types of change to be considered as learning. In the first two studies we compare Memory-Based and Error-Correction learning (MBL vs. ECL). MBL is championed by (computationally inclined) usage-based linguists; in it, change happens when new exemplar gets added to memory. ECL, however, formalizes change as filtering (in machine learning) or discrimination (in cognitive science), to minimize error in predicting an outcome. Our results show that ECL is a worthy opponent: it fits the experimental data better and has better biological (or cognitive) credibility.

In the next two studies, we take further steps. Naïve Discrimination Learning (NDL), our main computational modelling framework, makes use of certain *ad hoc* abstractions for input cues and outcomes in error-correction (discriminative) learning. We ask what would happen if those abstractions were linguistically informed: Langacker considers *units* as being "abstracted from usage events [...] through the reinforcement of recurring commonalities" (2019, p. 346). This is, indeed, suspiciously similar to a *relatively permanent change from experience*, once we resist the temptation to misinterpret usage-based *units* as static and idealized (an assumption usage-based linguists don't hold). With this in mind, we tested whether bottom-up cues, from the original NDL setup, can be coupled with and benefit from top-down, theoretically motivated cues, in learning the same lexical outcomes. The results, based on data from self-paced reading, show that both types of cues do discriminate lexical outcomes and contribute significantly to predicting reading latencies, although they have somewhat different roles. The results also show that learning is not limited to the actual linguistic cues focused on in the experiment; instead, learning applies to any and all cues present in the situation. Using insights from Reinforcement Learning we frame our findings in terms of exploration and exploitation.

I conclude with a simple point that the results our work, jointly taken, provide empirical traction for the theoretical point of Spreat & Spreat that "much like the law of gravity, the laws of learning are always in effect" (1982, p. 593).

The role of prediction errors in 4-year-olds' learning of English direct object datives.

Chiara Gambi, Cardiff University, gambic@cardiff.ac.uk
Kate Messenger, University of Warwick, K.Messenger@warwick.ac.uk

Is children's acquisition of structural knowledge driven by prediction errors? According to error-driven models of language acquisition (e.g., [1],[2]), children generate linguistic expectations about upcoming words, compare them to the linguistic input, and when they detect a mismatch (i.e., prediction error signal) they update their long-term linguistic knowledge. But we only have limited empirical evidence for this learning mechanism. Prediction error (induced by violations of verb-specific structural preferences) modulates the magnitude of both short-term [3] and cumulative priming effects [4] in production tasks using the English dative alternation. However, these studies tested 5 and 6 year olds, who are already able to understand and use both prepositional object (PO) datives and the more difficult direct object (DO) datives.

In contrast, we do not know whether younger children's acquisition of DOs is driven by prediction error. We test whether 4-year-olds' understanding of DOs improves more when children are exposed to input that encourages the generation of prediction error signals. Specifically, we contrast training conditions where the input allows children to generate strong expectations about upcoming words (which later turn out to be incorrect) to training conditions where the input does not support expectation-generation: Strong expectations, when disconfirmed, should lead to larger improvements in understanding.

We developed a novel web-based touchscreen task, comprising of three phases. In the first phase, we assessed children's baseline comprehension of DOs (pre-test, see Fig 1). Children listened to pre-recorded DO sentences whilst viewing pictures of the theme and recipient on a touchscreen, and then acted out their interpretation: Correct answers required dragging the theme picture (e.g., horse) towards the recipient picture (e.g., monkey). Then, we exposed children to one of four different training conditions, and finally assessed their DO comprehension skills again (post-test), using a different set of sentences.

During training, children (N = 98) listened to 12 dative sentences - either all POs or DOs (between participants). PO training conditions control for structural priming effects (exposure to DOs may increase children's post-test performance regardless of prediction error). Critically, we contrasted (also between participants) training sentences with an inanimate theme (e.g., frisbee in Fig 1) *versus* an animate theme (e.g., owl); the recipient was always animate (e.g., duck). Since inanimate referents are more likely to be themes, children looking at a frisbee and a duck could predict the frisbee would be the theme even before they heard the sentence [5]; but children looking at an owl and a duck could not make this prediction. Thus, after hearing *Winnie the Pooh will give…*, children exposed to inanimate themes should generate a stronger expectation for the theme (Pred condition), compared to children exposed to animate themes (NonPred). Importantly, this predictability manipulation should only affect learning for children trained on DOs because only these children had their expectations disconfirmed in the Pred condition by hearing the recipient before the theme (e.g., *Winnie the Pooh will give...the duck...the frisbee.*).

We assessed post-test performance (while controlling for pre-test scores) separately for test items with inanimate themes (AI) and those with animate themes (AA); AI items are easier for children [6], a finding we replicated (54.59% vs. 39.97% accurate). Interestingly, there was no structural priming effect, nor an interaction between sentence type (PO vs. DO) and predictability for the easier AI test trials (all p's > .360; see left-hand panel of Fig. 2). Importantly, however, for the more difficult AA test trials, we found not only a priming effect (B=0.65, SE=0.32, z=2.06, p=.039), but also a larger average improvement in comprehension accuracy (from pre- to post-test) for children exposed to DOs in the Pred condition (24.24%, N=22), compared to those exposed to DOs in the NonPred condition (<1%; N=20); improvement following PO training was low overall (see right panel, Fig. 2).

These findings provide preliminary evidence that prediction error drives acquisition of difficult direct object sentences in 4 year olds. However, the interaction between sentence type and predictability for AA test trials was only marginally significant (B=1.24, SE=0.63, z = 1.95, p =.051). We plan to resume data collection when the COVID-situation allows.

Figure 1. Schematic summary of the three-phase study design. For each of the three phases (pre-test, training, post-test) we show one representative item, with visual input at the top and sample sentences at the bottom. Pictures are screenshots from the web-based task, which children completed on a touchscreen tablet.

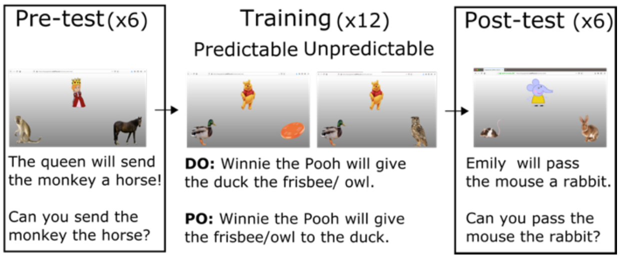
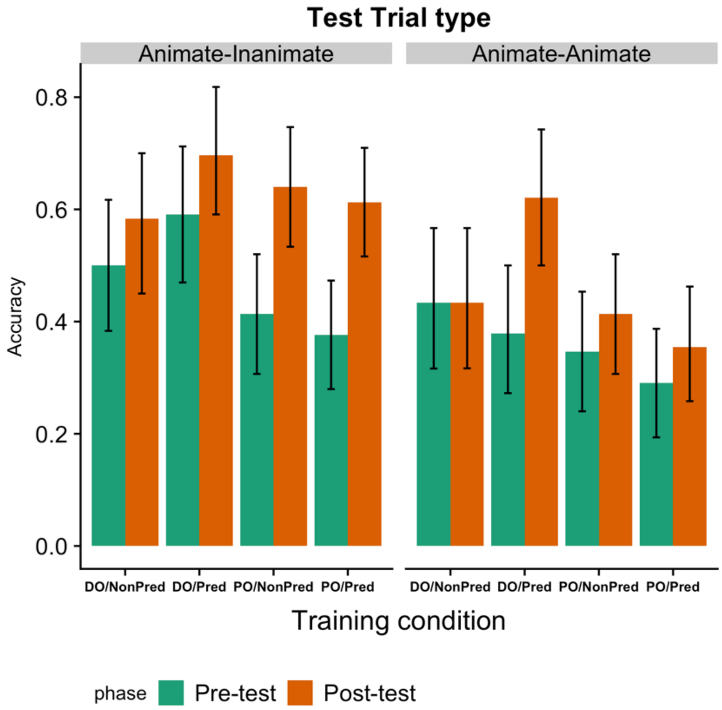


Figure 2. Mean pre-test (green bars) and post-test (orange bars) comprehension accuracy in the four training conditions (PO = prepositional object, DO = direct object, NonPred = unpredictable condition, Pred = predictable condition); the left panel shows data for the easier Animate-Inanimate (AI) test trials, while the right panel shows data for the more difficult Animate-Animate (AA) trials. Error bars represent 95% bootstrap CIs.



References

[1] Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14(2),* 179-211.

[2] Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review, 113(2),* 234-272.

[3] Peter, M., Chang, F., Pine, J. M., Blything, R., & Rowland, C. F. (2015). When and how do children develop knowledge of verb argument structure? Evidence from verb bias effects in a structural priming task. *Journal of Memory and Language, 81,* 1-15.

[4] Fazekas, J., Jessop, A., Pine, J. M. & Rowland, C. F. (in press). Do children learn from their predictions mistakes? A registered report evaluating error-based theories of language acquisition. *Royal Society Open Science*.

[5] Thothathiri, M. & Snedeker, J. Syntactic priming during language comprehension in three-and four-year-old children. *Journal of Memory and Language, 58(2),* 188-213.

[6] Buckle, L., Lieven, E. & Theakston, A.L. (2017). The effects of animacy and syntax on priming: A developmental study*. Frontiers in Psychology, 8*, doi:10.3389/fpsyg.2017.02246

Balancing information-structure and semantic constraints on construction choice: A discriminative learning model of passive and passive-like constructions in Mandarin Chinese (and Balinese and Hebrew).

Ben Ambridge

University of Liverpool, UK

ESRC International Centre for Language and Communicative Development (LuCiD)

Li Liu

Guangdong University of Foreign Studies, China

The goal of this study was to build a discriminative-learning model of how Mandarin speakers choose between one of four truth-value-identical constructions when producing a two-argument utterance, given two (often competing) constraints: (a) An information structure constraint which specifies that "The denotation of the *by*-phrase NP in a passive clause must denote something at least as new in the discourse as the subject". (Pullum 2014:64) and (b) A construction-semantic constraint such that the BEI Passive (1) and BA Active constructions (2), but not the Notional Passive (3) and SVO Active constructions (4), are associated with the meaning of affectedness of the PATIENT (i.e., of the OBJECT of the active forms).

	PATIENT Affected	PATIENT not (necessarily) affected
Topic = PATIENT	Mandarin O-BEI-SV passive (English OVS passive)	Mandarin OSV notional passive
Topic = AGENT	Mandarin S-BA-OV active	Mandarin SVO Active (English SVO Active)

(1) Lisi bei Zhangsan jiu le.
Lisi was saved by Zhangsan.

(2) Zhangsan ba Lisi jiu le.
Zhangsan saved Lisi.

(3) Zaofan Zhangsan chi le.
Zhangsan finished his breakfast.

(4) Zhangsan jiu le Lisi.
Zhangsan saved Lisi.

First, we conducted a grammaticality judgment study with 60 native speakers which confirmed that, across 57 verbs, semantic affectedness – as determined by a further 16 native speakers – determined each verb's relative acceptability in the BEI Passive and BA Active constructions, but not the Notional Passive and SVO Active constructions.

Second, in order to simulate acquisition of these competing constraints, we built a discriminative learning model that learns to map from corpus-derived input (information structure + verb semantics + lexical verb identity) to an output representation corresponding to these four constructions. The model was able to predict judgments of the relative acceptability of the test verbs in the BA Active and BEI Passive constructions, obtained in Study 1, with model-human correlations in the region of r=0.48 and r=0.33, respectively.

Third, in ongoing work, we are extending the model to simulate equivalent already-collected data for passive(-like) constructions in Balinese and Hebrew.

Together, these findings contribute to a growing body of evidence showing that error-driven-learning models in general, and Resorla-Wagner/Widrow-Hoff style discriminative learning models in particular, hold considerable promise as mechanistic accounts of language acquisition, extending this evidence to a new domain (verb argument structure) and to new languages.

Figure 1. Grammaticality Judgment scores (y axis) as a function of human semantic affectedness ratings (x axis)

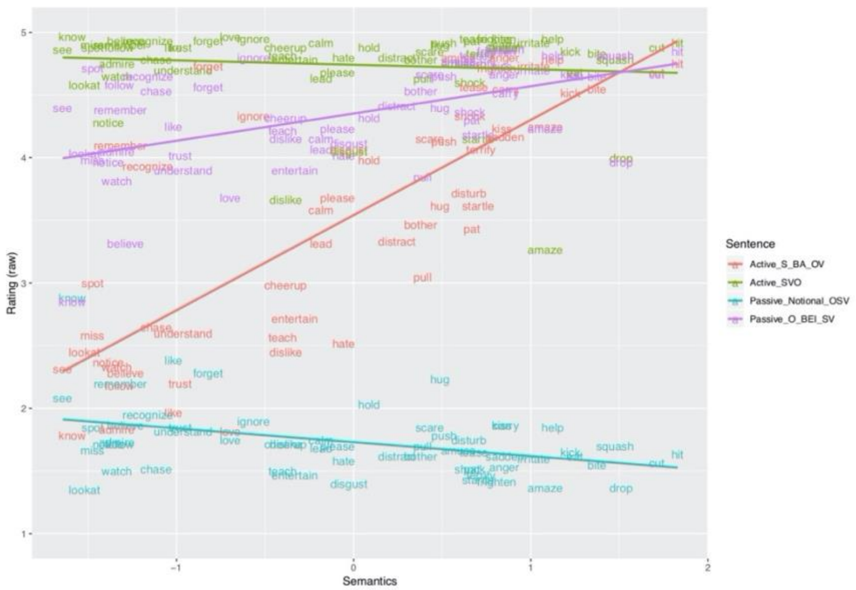


Figure 2. Architecture of the discriminative-learning model

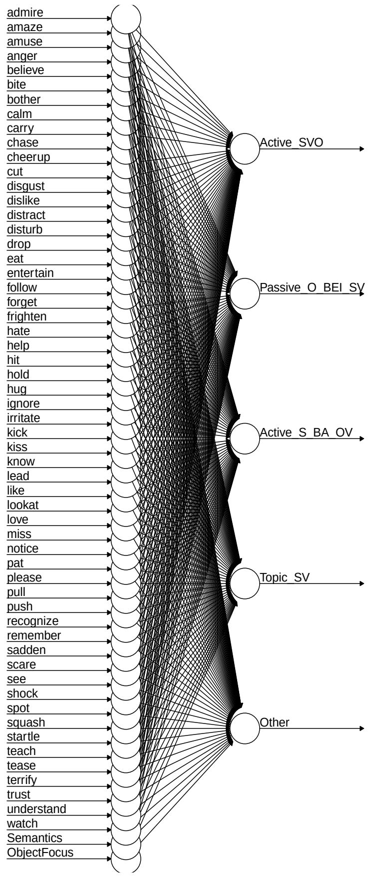
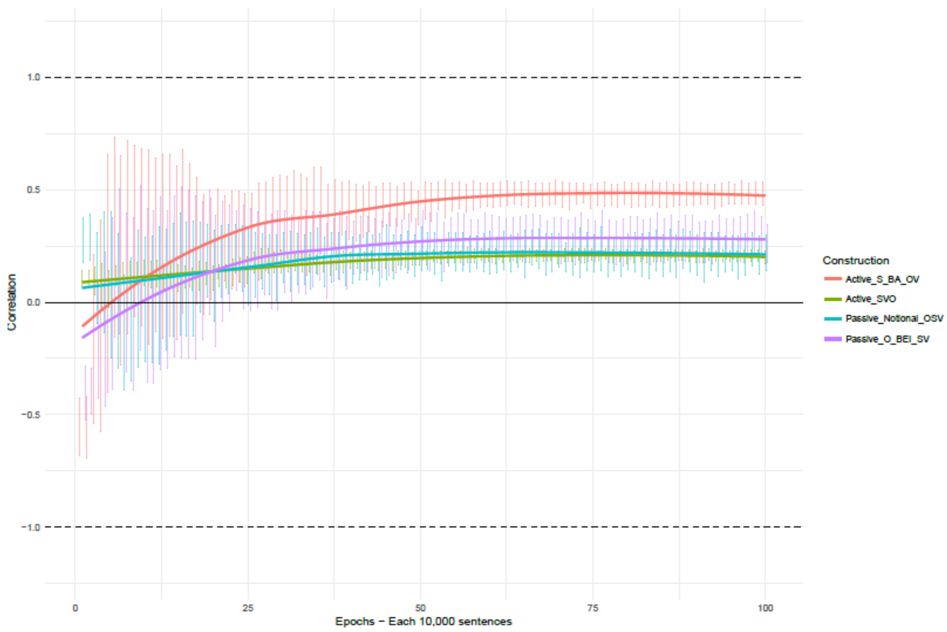


Figure 3. Model-human correlations



Can error-based models account for language processing via syntactic priming? Investigating the effects of task and learner characteristics

Marion Coumel[a], Ema Ushioda[b], Katherine Messenger[a]
[a]Psychology Department, University of Warwick, Coventry, UK
[b]Department of Applied Linguistics, University of Warwick, Coventry, UK

Recent psycholinguistic models propose that first (L1) and second language (L2) syntactic priming relies on an implicit, error-based language processing and learning mechanism[1,2]. Empirical support for this account mostly comes from studies obtaining L1 and L2 long-term priming[3,4] and inverse frequency effects[5,6] but some of the model's predictions remain largely unexplored. First, the model states that priming magnitude should vary with individuals' learning rate which should itself be determined by task characteristics. However, few studies have examined, for instance, whether task aspects such as the modality of prime sentences (i.e., auditory vs. visual) influences priming[7,8]. Second, few studies have tested the model's prediction that learner characteristics such as individual differences in attention and motivation should affect one's learning rate and thus, priming[9,10]. While most research testing these predictions targets L1 speakers, we expected L2 speakers' priming behaviour to be more sensitive to variation in task and learner characteristics given their overall reduced experience with the target language. Indeed, presenting prime sentences visually vs. auditorily might be particularly helpful for L2 speakers and attention and motivation are very relevant to second language learning[11,12]. Thus, the present study examined the effect of prime sentences' modality and individual differences in attention and motivation on L2 and L1 speakers' immediate and long-term syntactic priming.

Using an online written picture description task, we compared French L2 English speakers' and English L1 speakers' primed production of the active/passive syntactic alternation (Fig. 1). We manipulated between-subjects whether participants listened to (listening condition) or read the prime sentences (reading condition). We assessed attention (L2 and L1 speakers) and motivation (L2 speakers only) with questionnaires. We measured immediate priming (repeating a syntactic structure after a prime) and long-term priming (producing more target structures in immediate and delayed post-tests without primes relative to pre-tests).

Overall, we predicted that both speaker groups would show immediate and long-term priming and that higher attention levels (both groups) and higher motivational levels (L2 speakers) would lead to larger priming effects. However, because reading the primes (vs. listening to them) was expected to facilitate L2 speakers' processing of the target structures, we expected L2 speakers to show more immediate and long-term priming in the reading than in the listening condition. For the same reason, we predicted that higher attention and motivation levels would be more helpful in the listening than in the reading condition and thus, boost L2 speakers' priming more in the former than in the latter condition. On the contrary, we expected that L1 speakers would exhibit the same priming strength and that attention would have the same effect irrespective of modality conditions.

As predicted, both speaker groups experienced immediate and long-term priming in the immediate post-test (Fig. 2 & 3). However, only L2 speakers exhibited long-term priming in the delayed post-test. Regarding the effect of modality, the results support our predictions regarding L1 but not L2 speakers. Prime modality did not affect priming in either group. Moreover, we did not find any interaction between attention or motivation, prime modality and any of the three priming types in L2 speakers. Only L1 speakers who were more attentive to the stimuli and the task were also more likely to experience long-term priming in the immediate post-test across modality conditions.

Overall, the long-term priming effects provide evidence that syntactic priming is a language learning mechanism. The between-group difference in the delayed post-test is in line with the model's prediction that less experienced speakers should experience more learning. Yet, the findings do not support the predictions regarding the effect of modality and provide limited support for the predictions regarding the effect of learner characteristics.

References:
[1]Chang, Dell, & Bock, (2006), *Psychological Review*, *113*, 234-272
[2]Jackson, (2018), *Second Language Research, 34*, 539-552
[3]Hartsuiker and Kolk, (1998), *Language and Speech, 41*, 143-184
[4]Jackson & Ruf, (2017), *Applied Psycholinguistics, 38,* 315-345
[5]Hartsuiker & Westenberg, (2000), *Cognition, 75*(2), B27-39
[6]Kaan & Chun, (2017), *Language and Cognition,* 1-15
[7]Cleland & Pickering, (2006), *Journal of Memory and Language, 54*(2), 185-198
[8]Hartsuiker et al., (2008), *Journal of Memory and Language, 58*(2), 214-238
[9]Bock et al., (1992), *Psychological Review, 99*(1), 150-171
[10]Ivanova et al., (2020), *Journal of Memory and Language, 110*
[11]Takahashi, (2005), *Applied Linguistics, 26*, 90-120
[12]Ushioda, (2016), *Language Teaching, 49*, 564-577

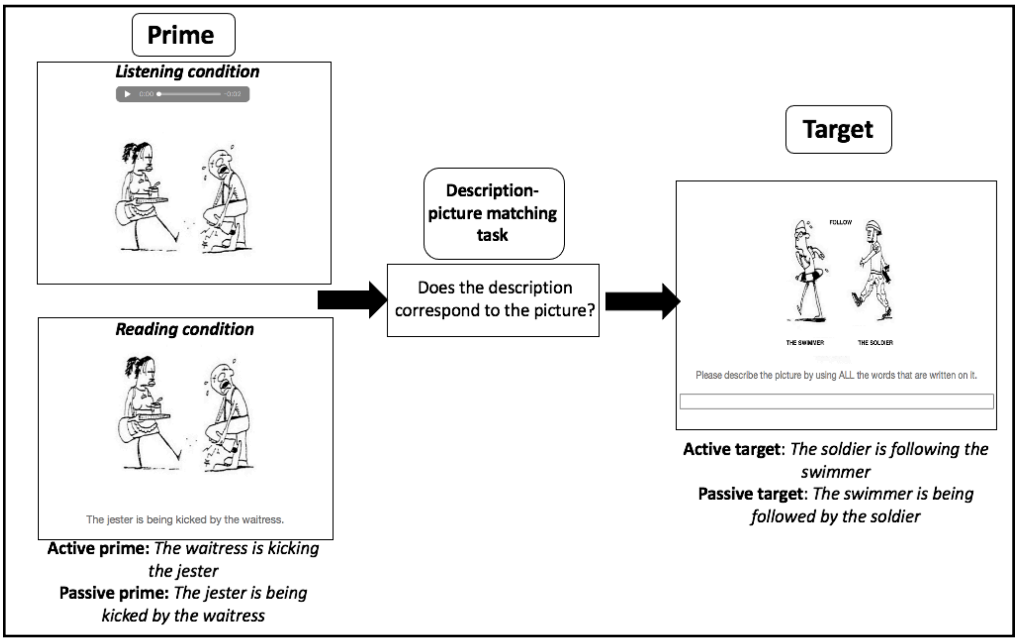


Figure 1. Experimental trial and example of active/ passive alternation.

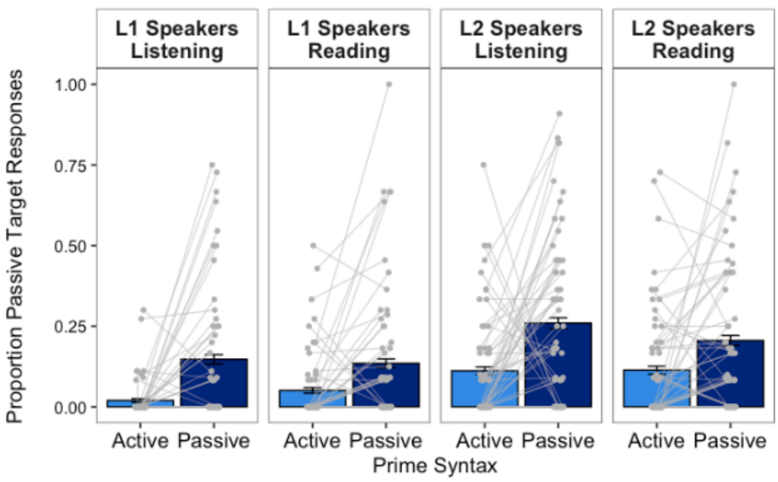


Figure 2. Passive responses in the immediate priming phase. Mean proportion of passive responses out of all transitive responses by prime syntax, prime modality and group condition in the immediate priming phase. Error bars indicate the standard error of the mean, grey dots indicate individual data points and grey lines individual priming effects.

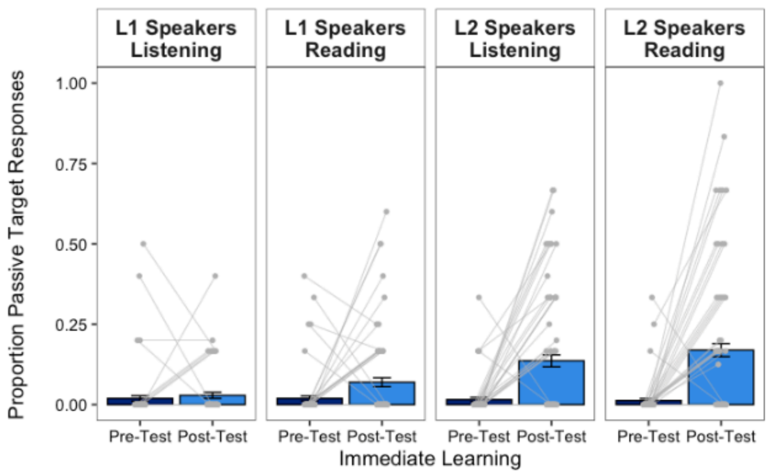


Figure 3. Passive responses in the pre- and immediate post-test. Mean proportion of passive descriptions in the pre- and immediate post-test out of all transitive descriptions by section, prime modality and group condition. Error bars indicate the standard error of the mean, grey dots indicate individual data points and grey lines individual priming effects.

Does learning occur in the absence of cues?

Sanne Poelstra[a], Jessie S. Nixon[b], Jacolien van Rij[a]
[a]University of Groningen, [b]University of Tübingen
s.poelstra.1@student.rug.nl, jessie.nixon@uni-tuebingen.de, j.c.van.rij@rug.nl

Discriminative, Error Driven Learning (EDL) is a theory and set of equations that model bottom-up learning by minimising the uncertainty in the learner's expectations about upcoming events. Well-known formalisations of EDL include the Rescorla-Wagner model (1972) and the almost identical Delta Rule (Widrow & Hoff, 1960). Generally, we model learning using a fully connected, two-layer network (i.e. input layer: cues; output layer: outcomes; no hidden layers). The informativeness of cues is a key notion in EDL: only if cues are present are the connection weights between cues and outcomes updated. With each learning event the connections between present cues and outcomes are strengthened, while the connections between present cues and absent outcomes are weakened.

However in their frequently cited paper, Van Hamme and Wasserman (1994) have argued based on experimental data, that we can also learn from absent cues. They proposed an adjustment to the Rescorla-Wagner model: An absent cue should be encoded negatively, which leads to a weakened connection between an absent cue and present outcome and a strengthened connection between an absent cue and an absent outcome.

In the present study we aim to disentangle these two models of EDL. We implemented two computational simulations that model the experimental study reported by Van Hamme and Wasserman (1994). One simulation implements the Rescorla-Wagner model; the second implements the adaptation proposed by Van Hamme and Wasserman, which allows for learning from absent cues. In this experiment, participants had to indicate how likely it was that certain foods caused an allergic reaction. There were three types of food, of which two occurred on each trial together with an outcome (an allergic reaction or not). The participants then estimated the causal relation on a scale from 0 to 8 for *all three foods*.

Figure 1 shows the results of our computational simulations. To model the rating scale, we calculated weights to *Allergic reaction* minus weights to *No reaction*. The simulations show that with the Van Hamme & Wasserman experiment design - specifically, when the response measure (rating) includes both outcomes (allergy, no reaction) - there are no substantial differences in weight development between the Rescorla-Wagner and the Van Hamme-Wasserman models. Although the strength of activations is numerically different, we do not have a link function sufficient to evaluate which model best describes the data. Therefore, the two models make essentially the same predictions. These simulations demonstrate that Van Hamme & Wasserman's experiment design was not able to tease apart which model performs better: so, whether or not we learn from absent cues remains an open question.

However, our simulations also showed that the two models do make different predictions during the later phases of the experiment – *if* the individual outcomes are tested separately (see Figure 2). When weights to *Allergic* are separated from weights to *No reaction*, the Rescorla-Wagner model (left) predicts that, for example, 'bran' continues to predict the allergic reaction; in contrast, by the end of Block 3, the Van Hamme-Wasserman model (right) predicts that 'bran' is a negative predictor of the allergic reaction.

Based on our simulations, in ongoing work, we are running a series of experiments, all modifications of Van Hamme and Wasserman's experiment, to test the predictions of the two model variants. We will test outcomes separately at the end of Block 3. In addition, it is not clear whether Van Hamme and Wasserman's experiment reflects implicit learning, because they explicitly measured participants' ratings of present and absent cues. However, we argue that EDL is an implicit process, which may be hindered by explicit inference. Therefore, we will also employ a forced-choice paradigm and a speeded response manipulation to test the effects of explicit reasoning vs. implicit error-driven learning.

References

Rescorla, R., & Wagner, A. (1972). A theory of pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical conditioning ii*, *64*, 99.

Van Hamme, L. J., & Wasserman, E. A. (1994). Cue competition in causality judgments: The role of nonpresentation of compound stimulus elements. *Learning and motivation*, *25*(2), 127–151.

Widrow, B., & Hoff, M. E. (1960). *Adaptive switching circuits* (Tech. Rep.). Stanford Univ Ca Stanford Electronics Labs.

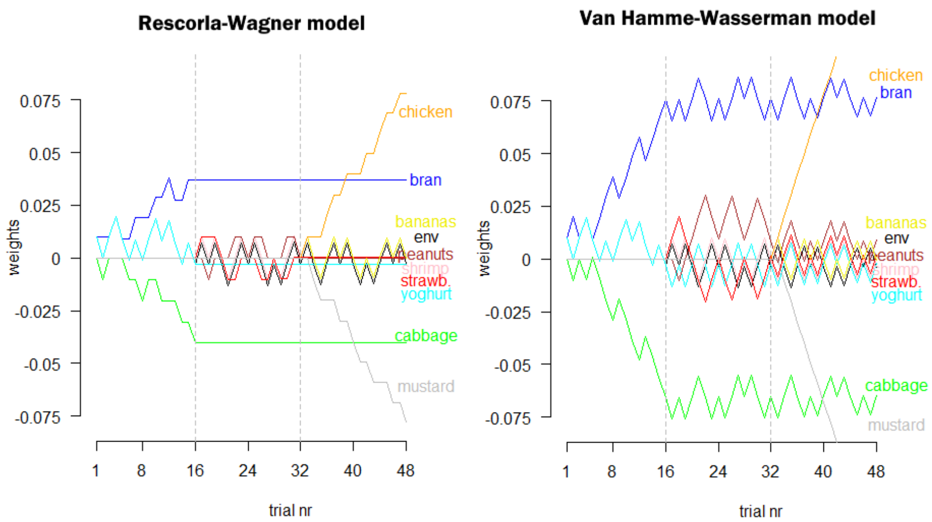


Figure 1: **Weights to *Allergic* minus the weights to *Not Allergic*** for each of the foods asked. Left: Rescorla-Wagner model. Right: Van Hamme-Wasserman model

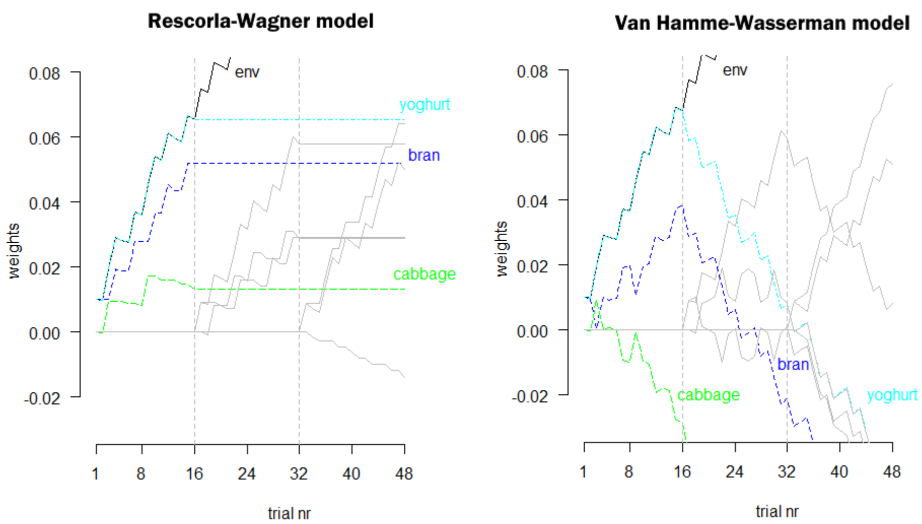


Figure 2: **Weights to *Allergic***. Left: Rescorla-Wagner model. Right: Van Hamme-Wasserman model

When backward transitional probabilities can be learned using forward prediction

Vsevolod Kapatsinski
University of Oregon

Backward transitional probability (BTP) is the probability of a word given the word that follows it, i.e., $p(\text{word}_i|\text{word}_{i+1})$ where *i* is position in an utterance. Perruchet and DeSaulty (2008) and Pellucchi et al. (2009) showed language learners to be capable of learning backward transitional probabilities from an input in which forward transitional probabilities, $p(\text{word}_i|\text{word}_{i-1})$, were controlled. These results have been argued to be problematic for predictive models of language learning in which learning results from predicting the future, and to provide decisive support for models that are capable of forming chunks based on either type of information (French et al., 2011; Perruchet, 2019). The present paper argues that this conclusion is premature: predictive models can become sensitive to either forward or backward transitional probabilities depending on a specific parameter setting.

While the discussion above has focused on recurrent networks, I trained a simpler two-layer network with an architecture previously proposed by Arnon and Ramscar (2012). The network was trained to predict the next word in the Switchboard Corpus (Godfrey et al., 1992) using its semantics and the identity of the preceding word, and used the Rescorla-Wagner learning rule (RW; Rescorla & Wagner, 1972). Semantic representations were either simple local codes or discretized Latent Semantic Analysis representations (Landauer & Dumais, 1997). In either case, they were more predictive of most of the words than the preceding word was. This greater predictiveness if crucial for obtaining the results below.

In its simplest form, RW updates cue→outcome associations based on the following two equations. For present outcomes and present cues, $\Delta w_{c\to o} = \alpha_c\beta_1(1-w)$; for absent outcomes and present cues, $\Delta w_{c\to o} = \alpha_c\beta_0(0-w)$. Crucially, the equations use different β parameters for present and absent outcomes. Figure 1 shows that if β_0 is much smaller than β_1(here it was set to zero and β_1 to .01 to show the extreme case), the item-to-item associations learned by the model reflect backward transitional probabilities and not forward ones. In contrast, if β_0 is set to the same value as β_1, the associations reflect forward transitional probabilities.

Since α and β are thought to reflect salience, a possible interpretation of this parameter manipulation is that the model's behavior depends on how much attention is allocated to absent forms. That is, sensitivity to backward transitional probability comes from paying little attention to absences.

This hypothesis generates a fresh perspective on differences between individuals and languages. It is often found that RW fits the data best if β_0 is smaller than β_1. Mckenzie and Mikkelsen (2007) have argued that this is because absences are less informative than presences, which suggests that the allocation of attention to absences may be adaptively adjusted by learners. A plausible mechanism for this adjustment is selection of attention allocation policies based on the prediction error that results from following a policy (Harmon et al., 2019). This explanation of sensitivity to BTP fits well with the finding that learners pick up on either BTP or FTP in an ambiguous artificial language based on which statistic is more informative in the participant's native language (Onnis & Thiessen, 2013).

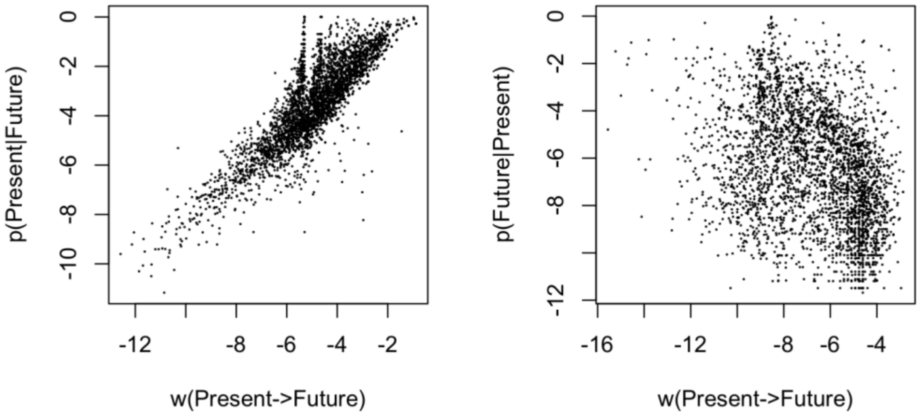


Figure 1. Forward associations (w) in a predictive model can track backward transitional probability (left panel, *r* = .76) rather than forward transitional probability (right panel, *r* = -.37) if there is cue competition between top-down and preceding-word cues to upcoming words, top-down cues are more predictive, and presences are much more salient than absences. Axes are log scaled. Semantic representations are localist. Points are individual tokens of words.

Arnon, I., & Ramscar, M. (2012). Granularity and the acquisition of grammatical gender: How order-of-acquisition affects what gets learned. *Cognition*, *122*(3), 292-305.

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*(4), 614-636.

Godfrey, J. J., Holliman, E. C., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. *IEEE International Conference on Acoustics, Speech, and Signal Processing* (Vol. 1, pp. 517-520).

Harmon, Z., Idemaru, K., & Kapatsinski, V. (2019). Learning mechanisms in cue reweighting. *Cognition*, *189*, 76-88.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211-240.

McKenzie, C. R., & Mikkelsen, L. A. (2007). A Bayesian view of covariation assessment. *Cognitive Psychology*, *54*(1), 33-61.

Onnis, L., & Thiessen, E. (2013). Language experience changes subsequent learning. *Cognition*, *126*(2), 268-284.

Pelucchi, B., Hay, J. F., & Saffran, J. R. (2009). Learning in reverse: Eight-month-old infants track backward transitional probabilities. *Cognition*, *113*(2), 244-247.

Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning.*Topics in Cognitive Science*, *11*(3), 520-35.

Perruchet, P., & Desaulty, S. (2008). A role for backward transitional probabilities in word segmentation? *Memory & Cognition*, *36*(7), 1299-1305.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp.64-99). New York: Appleton-Century-Crofts.

/LQHDU 'LVFULPLQDWLYH /HDUQLQJ LQ -XOLD
;XHIHQJ /XR <X <LQJ &KXDQJ DQG +DUDOG %DD\HQ

/LQHDU 'LVFULPLQDWLYH /HDUQLQJ LV D FRPSXWDWLRQDO IUDPHZRUN Z
PDNHV SUHGLFWLRQV DERXW ZRUG PHDQLQJV DQG IRUPV)RU
WR SUHGLFW ZRUG PHDQLQJV IURP ZRUG IRUPV)RU SURGXFW
ZRUG PHDQLQJV 7KH QHWZRUNV¶ ZHLJKWV DUH HVWLPDWHG X
PDWKHPDWLFV RI ZKLFK LV WKH VDPH DV WKDW RI PXOWLYDULD
PDWULFHV RI ZRUG IRUP DQG PHDQLQJ UHSUHVHQWDWLRQV
HTXLYDOHQW WR WKH FRPSUHKHQVLRQ DQG SURGXFWLRQ QH
HTXDWLRQV DQG = C UHVSHFWLYHO\ 2Q WKH FRPSUHKHQVLR
SUHGLFWV ZRUG PHDQLQJV VLPSO\ LQJ WKH LQSXW IRUP YHFWRU
SUHGLFWHG VHPDQWLF YHFWRUV E\ WDNLQJ WKH JROG ODEHO
VWURQJO\ FRUUHODWHG)RU SURGXFWLRQ WKH PRGHO ILUVW
s ZLWK %XW WKHQ LW SURFHHGV WR DVVHPEOH WKH SURQXQFLDW
DFFXUDF\ LV HYDOXDWHG E\ FRPSDULQJ WKH SUHGLFWHG ZRU
VWDQGDUG IRUPV
G.CDV LQLWLDOO\ LPSOHPHQWHG LQ 5 ZLWK WKH SDFNDJH
G/HLV H[WUHPHO\ WLPH DQG FRPSXWDWLRQDO SRZHU FRQVX
PRGHO RQ ODUJH GDWDVHWV LQ WKH SURJUDPPLQJ ODQJXDJH WKD
RSHUDWLRQV :H PDGH D FRPSOHWHO\ QHZ LPSOHPHQWDWLRQ
Cm/BGEZMLFK UHGXFHV FRPSXWDWLRQ WLPH DQG PHPRU\ GUDP
LPSOHPHQWHG &KROHVN\ GHFRPSRVLWLRQ ZKHQ FDOFXODWL
RI FDOFXODWLRQ DQG DFFXUDF\ DUH DOVR FRQVLGHUDEO\ LP
DOJRULWKPV IRU SURGXFWLRQ FRPSDUHG WR WKH RULJLQDO
E\qTKqBi?GQH DGGLWLRQ WR VSHHG XS PDWUL[RSHUDWLRQ ZH
IRUPDW LQ -XOLD WR GHDO ZLWK VSDUVH PDWULFHV
%HQFKPDUNLQJ VWXGLHV VKRZ WKDW LW VLJQLILFDQWO\ UHGXFHV FRPSXWDWL
PDNHV LW SRVVLEOH WR SURFHVV ODUJHU GDWDVHWV EXW D
DFFXUDFLHV)RU H[DPSOH IRU D GDWDVHW FRQWDLQLQJ
QRXQV WKH FURVV YDOLGDWLRQ DFFXUDF\ IRU SURGXFWLRQ
Cm/BGEZMLOH WKH DFFXUDF\ LQFUHDVHV IURP WR)RU
VKRUW)UHQFK XWWHUDQFHV ZLWK DX[LOLDU\ FRPSOH[LW\
FDQQRW SURFHVV LW EHFDXVH LW WDNHV WRR PXFK PHPRU\ +
KRXUV RI WUDLQLQJ DQG HYDOXDWLRQ RQ FURVV YDOLGDWLR
2WKHU IHDWXUHV LQFOXGH PXOWL WKUHDGLQJ FRPSXWLQJ DQG
+RII OHDUQLQJ UXOH LPSOHPHQWHG LQ WKH QH[W VWHS ZH DUH QRZ FR
LQJ LV WR GHVLJQ VHYHUDO KLJK OHYHO ZUDSSHU IXQFWLRQV
HDVH RI WKH XVHU 7KH IDVWHU DQG PRUH PHPRU\ HIILFLHQW LPSOHPHQWDWLRQ
KDV QRZ PDGH LW SRVVLEOH WR VWXG\ D ZLGHU UDQJH RI PR

5 + %DD\HQ HW DO ³7KH GLVFULPLQDWLYH OH[LFRQ $ XQLILHG FRPSXWD
SURFHVVLQJ LQ FRPSUHKHQVLRQ DQG SURGXFWLRQ JURXQGHG QRW LQ G
OHDUQLQJ´ &RPSOH[LW\ SS ± 2y,XRR88fkyRNf93N83NR
%HUQDUG :LGURZ DQG 0DUFLDQ (+RII ³$GDSWLYH VZLWFKLQJ FLUFXLWV
5HFRUG 3DUW ,9 SS ± 2y,XkRkjef /yk9R8jR

GDWDVHWV ZKLOH DW WKH VDPH WLPH PDNLQJ WKH PRGHO PR

Triphone meanings co-determine tongue shape during articulation: An ultrasound study

Motoki Saito, Fabian Tomaschek, R. Harald Baayen
Eberhard Karls Universität Tübingen

motoki.saito@, fabian.tomaschek@, harald.baayen@uni-tuebingen.de

Discriminative learning (Rescorla, 1988; Rescorla & Wagner, 1972) has been successfully employed to model a wide range of experimental data (e.g. Baayen, Milin, & Ramscar, 2016; Tomaschek, Plag, Ernestus, & Baayen, 2019). However, most of these studies have focused on the comprehension side of language processing (e.g. Baayen, Chuang, Shafaei-Bajestan, & Blevins, 2019; Shafaei-Bajestan & Baayen, 2018) or on the speech signal produced (Tomaschek et al., 2019). What is not well understood at present is to what extent words' meanings help shape the way in which words are articulated.

Saito (2020a) begins to address this question by investigating word-final triphones that contained the stem vowel [a(ː)] and the word-final [t]. These authors observed lower tongue tip and higher tongue body positions for word-final triphones that were more strongly implicated in the mappings between form and meaning. Triphones' degree of semantic support was estimated using a Linear Discriminative Learning (LDL) (Baayen et al., 2019) model; the degree of semantic support can be understood as an operationalization of the classical concept of *functional load* (Saito, 2020a). The tongue movement data in their study was recorded with electromagnetic articulography (EMA), where only a few sensors on the tongue can be tracked. To consolidate this effect of functional load on articulation, the present study made use of ultrasound as experimental method, instead of EMA. Ultrasound offers the possibility to study the movement of much larger parts of the tongue, especially when analyzed with GAMMs (Saito, 2020b).

The ultrasound data in the present study consist of 20 participants articulating 126 German inflected verbs with the stem vowel [aː]. These verbs were combined with two types of pronouns ([ziː] vs [(v)ːr]) and the two types of suffixes ([t] vs [n]). The verbs were selected subject to the criterion that at most one segment intervened between the stem vowel and the suffix (e.g. *malt* [maːlt]).

Functional load, the semantic contribution of sublexical units (triphones) to the target meaning, was first pitted against word frequency and three commonly used measures from Naive Discriminative Learning (NDL) (Tomaschek et al., 2019), i.e. activation, prior, and activation diversity, in order to replicate the previous finding that functional load outperforms frequency (Saito, 2020a) and the simple NDL measures (Baayen et al., 2019). To this end, a Random Forest model was fitted with brightness values in the ultrasound image as response variable. The functional load of the word-final triphone emerged with the highest variable importance.

A Generalized Additive Mixed-effects Model (GAMM) (Wood, 2006) was then fitted to model ultrasound images with x- and y-coordinate values and functional load as predictors. The ultrasound images fitted with the GAM model are shown in Figure 1 for a high (quantile=0.9) and a low (quantile=0.1) value of the functional load of the word-final triphone. As expected and consistent with Saito (2020a), the tongue body is positioned higher for the high functional load (in the leftmost plot), compared to a low functional load (second plot). Conversely, the tongue tip is higher in the second plot than in the first plot. The differences between these two plots are presented in the third plot. The rightmost panel highlights where differences are significant.

The present finding indicates that a greater functional load of the word-final triphone induces a more bulged shape of the tongue. This result provides further support for the hypothesis that semantics influences fine details of phonetic realizations (Saito, 2020a) and challenges classical views of speech production such as WEAVER++ (e.g. Levelt, Roelofs, & Meyer, 1999; Levelt & Wheeldon, 1994).

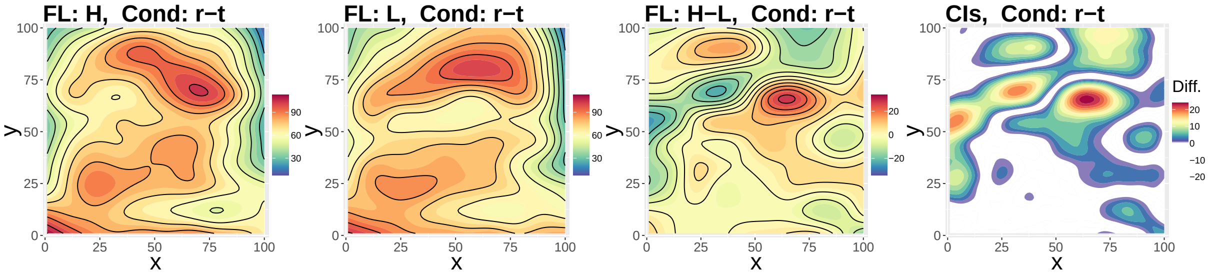


Figure 1: Fitted ultrasound images with GAM. The mouth front is to the right of each image. Warmer colors represent higher values and colder colors represent lower values. The first (leftmost) and second plots are when the functional load is high and low for each. The third plot shows how the two surfaces differ. The rightmost visualizes where the two surfaces differ significantly. Warmer colors indicate larger differences between confidence regions.

References

Baayen, R. H., Chuang, Y.-Y., Shafaei-Bajestan, E., & Blevins, J. P. (2019). The Discriminative Lexicon: A Unified Computational Model for the Lexicon and Lexical Processing in Comprehension and Production Grounded Not in (De)Composition but in Linear Discriminative Learning. *Complexity*, 1–39. doi: 10.1155/2019/4895891

Baayen, R. H., Milin, P., & Ramscar, M. (2016). Frequency in lexical processing. *Aphasiology*, *30*(11), 1174–1220. doi: 10.1080/02687038.2016.1147767

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–75.

Levelt, W. J. M., & Wheeldon, L. (1994). Do speakers have access to a mental syllabary? *Cognition*, *50*, 239–269.

Rescorla, R. A. (1988). Pavlovian Conditioning: It's Not What You Think It Is. *American Psychologist*, *43*(3), 151–160.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning ii: Current research and theory* (pp. 64–99). New York: Appleton-Century-Crofts.

Saito, M. (2020a). Relative functional load determines co-articulatory movements of the tongue tip. In *Proceedings of issp 2020 - 12th international seminar on speech production.*

Saito, M. (2020b). An ultrasound study of frequency and co-articulation. In *Proceedings of issp 2020 - 12th international seminar on speech production.*

Shafaei-Bajestan, E., & Baayen, R. H. (2018). Wide learning for auditory comprehension. In *Proceedings of the annual conference of the international speech communication association, interspeech* (pp. 966–970). doi: 10.21437/Interspeech.2018-2420

Tomaschek, F., Plag, I., Ernestus, M., & Baayen, R. H. (2019). Phonetic effects of morphology and context: Modeling the duration of word-final S in English with naïve discriminative learning. *Journal of Linguistics*, 1–39. doi: 10.1017/S0022226719000203

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Boca Raton, Florida, U.S.A.: CRC Press.