A hierarchical Bayesian model for syntactic priming

Weijie Xu (weijie.xu@uci.edu)
Department of Language Science, 3151 Social Science Plaza
Irvine, CA 92697 USA

Richard Futrell (rfutrell@uci.edu)
Department of Language Science, 3151 Social Science Plaza
Irvine, CA 92697 USA

Abstract

The effect of syntactic priming exhibits three well-documented empirical properties: the lexical boost, the inverse frequency effect, and the asymmetrical decay. We aim to show how these three empirical phenomena can be reconciled in a general learning framework, the hierarchical Bayesian model (HBM). The model represents syntactic knowledge in a hierarchical structure of syntactic statistics, where a lower level represents the verb-specific biases of syntactic decisions, and a higher level represents the abstract bias as an aggregation of verb-specific biases. This knowledge is updated in response to experience by Bayesian inference. In simulations, we show that the HBM captures the above-mentioned properties of syntactic priming. The results indicate that some properties of priming which are usually explained by a residual activation account can also be explained by an implicit learning account. We also discuss the model's implications for the lexical basis of syntactic priming.

Keywords: syntactic priming; hierarchical Bayesian model; lexical boost; inverse frequency; asymmetrical decay; implicit learning; lexical-based priming

Introduction

Syntactic priming refers to an adaptive behavior where exposure to a specific syntactic structure influences a language user's subsequent processing of similar structures. From the production side, speakers tend to repeat the syntactic structure they have recently been exposed to (Bock, 1986; Pickering & Ferreira, 2008; Bock & Loebell, 1990; Branigan, Pickering, & Cleland, 2000). For example, after a person produces or perceives a sentence such as "the boy gave the girl the ball" (called the **prime**), there is a reliable increase in the probability that their description of a subsequent ditransitive event (called the **target**) will use the same syntactic form, in this case using the double-object construction (DO; as in "give *X Y*") rather than a prepositional-object construction (PO; as in "give *X* to *Y*").

Empirical Background

A large body of work has examined factors that modulate the strength of syntactic priming (Mahowald, James, Futrell, & Gibson, 2016). In (1) we summarize three major findings from this literature, and review them below.

(1) Three empirical properties of syntactic priming

    a. *Lexical Boost*: Lexical alignment, especially verb overlap, between the prime and the target enhances the priming effect size.

    b. *Inverse Frequency*: Less frequent prime constructions elicit stronger priming effect.

    c. *Asymmetrical Decay*: Lexical boost decays faster than the abstract verb-independent priming effect.

The **lexical boost** refers to an increase in the effect size of syntactic priming when the prime sentence and the target sentence share critical lexical items (Pickering & Branigan, 1998; Gries, 2005). For example, if both the prime and the target use the same verb "give", then priming of the DO construction is stronger than if the prime and target involved different verbs. In comprehension, the alignment of verb lexical items has been found to be necessary to observe a reliable priming effect (Tooley & Traxler, 2010).

Another well-documented property of syntactic priming is the **inverse frequency effect**, where constructions that are less frequent in language experience are susceptible to a stronger priming effect (Scheepers, 2003; Kaschak, Kutta, & Jones, 2011; Hartsuiker & Kolk, 1998; Bock, 1986; Ferreira, 2003). For example, in English usage, PO constructions are overall rarer than DO; consequently, a PO prime causes a larger increase in the probability of a PO target than a DO prime does for a DO target. A more precise characterization of the inverse frequency effect has been put forward by Jaeger and Snider (2013), who claim that the effect size of priming is directly proportional to the prediction error during processing, which in turn depends on statistical regularities in the language experience. In that case, infrequent constructions or co-occurrences cause larger prediction errors and thus stronger priming effect.

Our last phenomenon of interest has to do with the decay of the size of the priming effect as a function of time between the prime and target. Decay of syntactic priming has been observed in both the short-term (Levelt & Kelter, 1982; Branigan, Pickering, & Cleland, 1999) and the long-term (Bock, Dell, Chang, & Onishi, 2007; Bock & Griffin, 2000; Kaschak, Kutta, & Schatschneider, 2011), in a way that depends on lexical alignment between the prime and target. In particular, **asymmetrical decay** refers to the fact that the lexical boost decays more rapidly than the abstract priming effect, where there is no lexical alignment (Bock & Griffin, 2000; Hartsuiker, Bernolet, Schoonbaert, Speybroeck, & Vanderelst, 2008).

Related Work

Two lines of theories are proposed in the literature, the residual activation account and the implicit learning account. Each account captures some but not all the empirical properties mentioned above.

The **residual activation** account of priming holds that production or comprehension of a syntactic form is influenced by its level of *activation*. Upon encountering a sentence, activation is boosted for the representation of both the lexical items and the syntactic structure, facilitating future reuse (Pickering & Branigan, 1998). Based on a spreading activation model of lemma retrieval (Roelofs, 1992), Pickering and Branigan (1998) specify a representational architecture where nodes representing verb lemmas are connected to the combinatorial nodes representing syntactic structures. In this model, a boost in activation on a certain verb lemma spreads outward to the combinatorial nodes and to other verb lemmas, which in turn further contribute to the activation of the combinatorial nodes. This residual activation account predicts the the lexical boost effect, since the the verb lemma used in the prime is directly activated and receives the strongest activation boost. However, the residual activation account has difficulty explaining long-term priming, since all activation is posited to decay over time.

The **implicit learning** account proposes that syntactic priming is a kind of syntactic learning. This idea was originally implemented in a connectionist model that predicts the next word based on the previously heard word (Chang, Dell, & Bock, 2006). In this model, prediction error serves as an explicit learning signal, based on which the weights between nodes are updated via backpropagation. The implicit learning account predicts that the priming effect size is proportional to the prediction error when processing the prime, thus accounting for the inverse frequency effect (Jaeger & Snider, 2013). This error-based effect is also reproduced in learning models that do not include an explicit error signal, such as the Bayesian belief-update (Kleinschmidt, Fine, & Jaeger, 2012; Fine, Jaeger, Farmer, & Qian, 2013; Gershman, 2015). However, compared to the residual activation account, the implicit learning account struggles to capture the observation that the syntactic priming effect can sometimes be short-lived.

Since neither account simultaneously explains all the three properties in (1), some studies propose a hybrid account, holding that both residual activation and implicit learning play a role. This amalgamation is also motivated by the observation of asymmetrical decay. In particular, residual activation explains the rapid decay of lexical boost, whereas long-lived abstract priming is mostly driven by implicit learning (Hartsuiker et al., 2008; Bock & Griffin, 2000). In line with this hybrid mechanism, Reitter, Keller, and Moore (2011) propose an activation-based model in the ACT-R framework (Lewis & Vasishth, 2005), where lexical chunks and syntactic chunks are represented separately in the declarative memory. The priming effect results both from an increase of the base-level activation of the syntactic chunk and from the spread-

ing activation from lexical chunks. Similar to Pickering and Branigan (1998), the spreading activation mechanism allows the model to capture lexical boost. The model also captures the inverse frequency effect because of the diminished base-level learning for high-frequency chunks. Moreover, the model captures the asymmetrical decay since the spreading activation in ACT-R only goes from chunks that are temporarily held in the working memory buffer to chunks in long-term memory. Therefore, when lexical chunks are evacuated from working memory, they cannot spread activation anymore.

The Current Study

We develop a computational-level theory of syntactic priming in terms of a hierarchical Bayesian model (HBM) of syntactic knowledge. Our HBM extends the basic Bayesian belief-update model by representing syntactic knowledge at multiple levels of abstraction (Kemp, Perfors, & Tenenbaum, 2007). As a general learning framework, HBMs have been broadly applied to speech perception (Kleinschmidt & Jaeger, 2015), the emergence of communicative systems (Hawkins et al., 2023), and human cognition in general (Tenenbaum, Kemp, Griffiths, & Goodman, 2011). We show how this general learning framework can capture the three empirical properties of priming in (1), namely the lexical boost effect, the inverse frequency effect, and asymmetrical decay.

Modeling Framework

Representing Syntactic Knowledge

The core of the priming effect is a decision-making problem: the decision between multiple syntactic frames for the same message in production, or between multiple parses for an ambiguous input in comprehension. The decision-making can be characterized by a probability distribution over the syntactic alternatives of interest. In the production of ditransitives, for instance, the decision between dative object (DO) and prepositional object (PO) can be captured by a Bernoulli distribution, whose parameter represents the production probability for DO versus PO. We aim to characterize this probability distribution and how it changes in response to experience.

Following previous work, we represent a person's syntactic knowledge as syntactic statistics: counts of how often one expects to experience different syntactic constructions (Fine et al., 2013; Jaeger & Snider, 2013), represented at different levels of abstraction. We posit a hierarchical structure for syntactic statistics, as illustrated in Figure 1. We specify two levels of abstraction. The lower level captures verb-specific statistics ϕ_v, such that each verb imposes its own bias (Bernolet & Hartsuiker, 2010). The verb-specific statistics echo the lexicalist view of sentence processing, where syntactic knowledge is accessible via lexical representations (MacDonald, Pearlmutter, & Seidenberg, 1994; McRae, Spivey-Knowlton, & Tanenhaus, 1998; Spivey-Knowlton & Sedivy, 1995). The higher-level abstraction Θ is an aggregation over all the lower-level statistics and represents the general decision bias across all the verbs.

Abstract Level of General Bias

Lexical Level of Verb-Specific Bias

$$\Theta$$
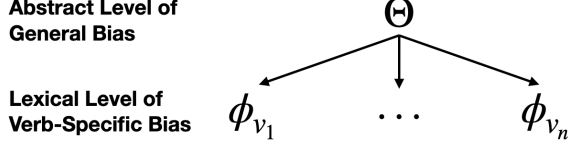
$$\phi_{v_1} \quad \cdots \quad \phi_{v_n}$$

Figure 1: Hierarchical representation of syntactic statistics.

Taking ditransitive priming as an example, the variables in Figure 1 would represent counts over DO constructions and PO constructions, conditional on specific verbs (the ϕ_v) or aggregating across verbs (the high-level Θ). However, this hierarchical structure of syntactic statistics is not limited to any specific construction (e.g. ditransitives). The idea can be applied to any scenario that includes processing decisions, such as ambiguity resolution (for example, garden-path sentences) in comprehension and the choice of syntactic frames in production (for example, passive vs. active for transitives).

Updating Syntactic Knowledge

We posit that syntactic learning takes the form of Bayesian updates to the hierarchical model in Figure 1. We explain how this works first using a familiar non-hierarchical Beta-Binomial model, and then the full hierarchical model.

Beta-Binomial Process. The Beta-Binomial model is a simple version of Bayesian belief updating. It postulates a generative model where a set of decisions x between two outcomes are sampled from a Binomial distribution parameterized by variable Θ, which in turn is sampled from a prior Beta distribution with hyperparameters $\alpha, \beta > 0$:

$$\Theta \sim \text{Beta}(\alpha, \beta) \tag{1}$$

$$x \sim \text{Binomial}(\Theta). \tag{2}$$

Applied to syntactic priming, the variable Θ represents the probability of the DO form versus the PO form, and the hyperparameters α, β in the Beta distribution represent prior expectations for frequencies of the DO and PO outcome respectively. After observing an outcome x, the variable Θ updates by Bayesian inference, following the new distribution

$$p(\Theta \mid x) \propto p(x \mid \Theta)p(\Theta). \tag{3}$$

This updated distribution takes a simple form. If one observes N datapoints containing x examples of a DO outcome, the updated decision probabilities follow

$$\Theta \mid x \sim \text{Beta}(\alpha + x, \beta + N - x). \tag{4}$$

Thus observing many DO outcomes will increase the probability of future DO outcomes. Having demonstrated the logic of Bayesian belief updating in a simple setting, we now turn to our hierarchical model.

Hierarchical Bayesian Model. As an extension of the Beta-Binomial process, we specify a hierarchical structure for syntactic statistics. Instead of having a single parameter Θ giving the decision bias, we posit that each verb has an individual decision bias ϕ_v, which is in turn sampled from a global decision bias variable Θ that ties the individual verb biases together. Specifically, we posit a generative model

$$\Theta \sim \text{Beta}(1, 1) \tag{5}$$

$$\phi_v \sim \text{Beta}(\alpha\Theta, \alpha(1 - \Theta)) \tag{6}$$

$$x_v \sim \text{Binomial}(\phi_v). \tag{7}$$

In this generative model, a global decision bias variable Θ is sampled from a distribution $\text{Beta}(1, 1)$ (which is uniform on the interval $[0, 1]$), then for each verb v, an individual decision bias ϕ_v is sampled from a Beta distribution parameterized by Θ. Here the parameter $\alpha > 0$ represents the relative importance of the global prior Θ in determining the verb-specific biases ϕ_v.

The hierarchical structure of the generative model means that experience with one verb can cause changes not only to that verb's decision bias, but also to the global decision bias, which in turn affects other verbs. Given counts of x_v DO outcomes for a specific verb v, when we are considering the update to the statistics for the currently-observed verb v, then the update is given by Bayes' rule applied to ϕ_v alone:

$$p(\phi_v \mid x_v) \propto p(x_v \mid \phi_v)p(\phi_v). \tag{8}$$

When we are considering the update to the variable ϕ_w for a *different* verb $w \neq v$, then the update is mediated through the updated global statistics Θ:

$$p(\phi_w \mid x_v) = \int p(\phi_w \mid \Theta)p(\Theta \mid x_v)\, \text{d}\Theta. \tag{9}$$

Therefore, information flows both bottom-up and top-down in the HBM. When the new data x_v for verb v is encountered, it impacts the verb-specific parameter ϕ_v and the effect goes bottom up to impact the higher-level Θ. Since this abstract Θ governs and constrain all the verb-specific parameters, as a top-down process, the effect on Θ in turn influences other ϕ_w. In this way, learning from the data of one verb can be generalized to others. We implement the learning of HBM using `WebPPL` (Goodman & Stuhlmüller, 2014).

Model Evaluation

In two simulations, we evaluate our model on the experimental materials of Pickering and Branigan (1998)'s Experiment 1, aiming to capture the three empirical properties of syntactic priming outlined earlier. The original experiment is in a trial-to-trial production priming paradigm using English ditransitives. There are 32 items, each consisting of a prime and a target as in (2) below. At the prime, participants are presented with the partial input of a ditransitive sentence until the first post-verbal object noun phrase, and are asked to complete the sentence. At the target, participants are asked to complete another partial sentence input that includes a subject noun phrase and a ditransitive verb. The prime sentences
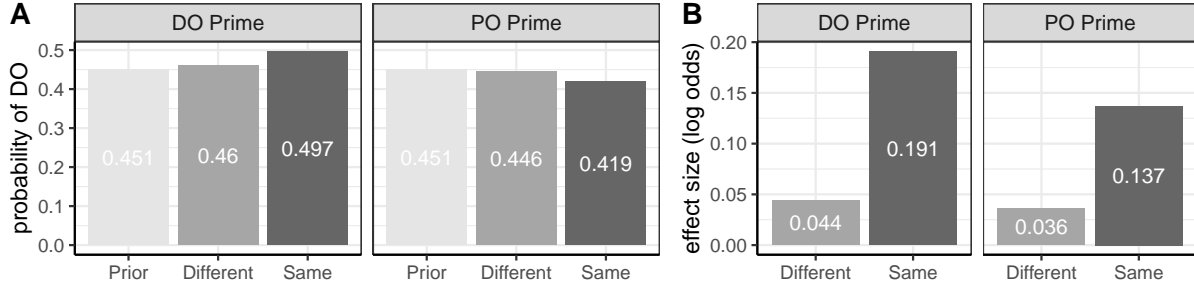
Figure 2: Results of Simulation 1. *Panel A*: Model-estimated average prior and posterior probability of DO for the target verbs in Pickering and Branigan (1998); `Same` refers to the condition with verb overlap between the prime and the target; `Different` refers to the condition without verb overlap. *Panel B*: Model-predicted priming effect size, calculated as the difference of log-odds between the posterior and the prior for DO.

are manipulated in a 2×2 design. First, the prime is biased towards either DO or PO structure. Second, verb overlap is manipulated such that the prime either shares or does not share the same verb as the target.

(2) Sample stimuli from Pickering and Branigan (1998)

    a. *Prime*:
      The racing driver showed the torn overall...
      The racing driver showed the helpful mechanic...
      The racing driver gave the torn overall...
      The racing driver gave the helpful mechanic...
    b. *Target*: The patient showed...

Simulation 1: Lexical Boost and Inverse Frequency

This first simulation aims to capture both the lexical boost and the inverse frequency effect. First, we expect the materials with verb overlap in Pickering and Branigan (1998) to generate a stronger priming effect in the model output, compared to the condition without verb overlap. Second, we expect the model to generate stronger priming effect for the ditransitive structure that is favored *a priori* before priming.

Estimating the prior. We first construct a dataset to represent the participant's language experience prior to the priming experiment.[1] As summarized in Table 1, this dataset contains 100 datapoints distributed across the nine verbs used in the original experiment. Each data point represents an instance of verb use in one of the two ditransitive structures (e.g. `<give:DO>`, `<show:PO>`). The frequency of each verb and their relative use of DO/PO are specified based on counts from the British National Corpus (Zhou & Frank, 2023; Yi, Koenig, & Roland, 2019).[2] The model starts with the global bias Θ and verb-specific biases ϕ_v derived from this dataset

---

[1]In Bayesian learning, this prior dataset corresponds to pseudocounts. The larger the size of the prior data, the smaller the priming effect for the same amount of exposure data.

[2]The verb frequencies in Table 1 are sampled from a multinomial distribution with the parameter $N = 100$ and the parameter **p** corresponding to the verb frequencies in the corpus. The DO/PO frequency for each verb is generated in a similar way but from a binomial distribution.

following the procedures above. This initial state is the prior distribution which will be compared with the updated posterior distribution after exposure to the priming data.

Verb	DO Freq.	PO Freq.	Verb Freq.
give	51	20	71
show	1	3	4
send	5	8	13
lend	1	0	1
hand	0	3	3
loan	0	0	0
offer	2	4	6
sell	0	2	2
post	0	0	0
in total	60	40	100

Table 1: Verb frequencies and counts of DO/PO ditransitives used to form the model's prior distribution.

Inferring the posterior. Starting with the prior defined above, the model infers a posterior based on the exposure data presented in the primes in Pickering and Branigan (1998). The posterior learned by the model is a joint distribution for both the global and verb-specific decision biases. However, the parameters that can be directly measured from the behavioral data are the verb-specific parameters, that is, the proportion of DO/PO use associated with the specific verbs in the target sentence of each item. Therefore, for the target verb in each item, we obtain the verb-specific posterior through marginalization as given in Eq. 9.

Result. Figure 2A shows the average DO probability for the target verbs in each condition. The control condition shows the prior DO probability before priming. Although there are more DOs in the prior data based on the raw frequency in Table 1, the model infers a global bias slightly against DO (i.e. $p(\text{DO}) < 0.5$), because the DO preference in the raw frequency is mainly driven by one single frequent verb *give* in the prior dataset. After updating based on the prime data, we

find a priming effect in the form of an increase in DO probability with DO primes and a decrease in DO probability with PO primes. Figure 2B shows the priming effect size more directly by calculating the difference of log-odds between the prior and the posterior of DO for each target verb.

Consistent with lexical boost, the model predicts a stronger priming effect when the prime and the target share the same verb. Consistent with the inverse frequency effect, the model predicts that DO primes—which is the structure less preferred in the prior—should elicit a stronger priming effect.

Discussion. The model's ability to capture the inverse frequency effect comes with Bayesian learning in general, where events that are more surprising induce larger updates (Courville, Daw, & Touretzky, 2006). More crucially, the model is also able to capture lexical boost. In fact, in the Bayesian framework, a better way to interpret the effect of verb (non-)overlap is *lexical transfer*, where the learning effect on one verb is generalized to others. It is worth noting that although our model successfully captures the qualitative conclusions in Pickering and Branigan (1998), it is difficult to directly compare the model-predicted results with the empirical data, because experimental work rarely reports the baseline pre-priming DO probability.

Simulation 2: Asymmetrical Decay

In Simulation 2, we aim to model the asymmetrical decay in syntactic priming, where lexical boost decays more rapidly than the abstract priming effect without verb overlap. The prior estimation follows the procedure identical to Simulation 1, using the prior data in Table 1.

The simulation is conducted in two groups, an exposure and a control group. For the exposure group, the model is first conditioned on the prime sentence. For the control group, there is no conditioning on prime sentences. Then, both groups are exposed to two batches of additional data, each containing 100 samples from the frequency distribution in Table 1. These post-priming data reflect the average effect of post-priming trials and experience. The priming effect size is then calculated as the difference of log-odds for DO between the model-inferred posterior in the exposure group and the one in the control group.

Result. Figure 3 shows the model-predicted priming effect size as a function of the number of batches of post-priming data. The result with zero batch of post-priming data is the same as the result in Simulation 1 (Figure 2B), and is repeated here for the convenience of comparison. First, as in Simulation 1, the model reproduces the lexical boost and the inverse frequency effect with additional post-priming data. Second, the result indicates a decay of the priming effect in general for both conditions of verb overlap: The priming effect size decreases as more and more post-priming data are taken by the model. Critically, the model also demonstrates an asymmetry of decay: The priming effect size diminishes more than twice as rapidly in the condition with verb overlap compared to the
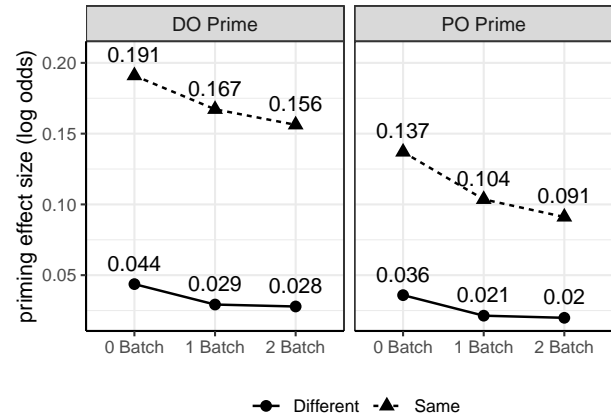

Figure 3: Simulation 2 model-predicted priming effect size as a function of the number of additional batches of post-priming data. Effect size calculated as in Simulation 1.

condition without verb overlap.

Discussion. As expected, our model successfully captures the asymmetrical longevity of syntactic priming, such that the lexical boost effect decays more rapidly than the priming effect without verb overlap. To our knowledge, the decay of syntactic priming and the distinction between long- and short-term priming has rarely been explained from the perspective of implicit learning, and has actually been considered one of the major limitations of the implicit learning account. However, the result of the current simulation demonstrates that at least some aspects of the long-/short-term distinction in priming can be captured by the implicit learning account in principle. We discuss the potential mechanism of how the HBM can capture the asymmetrical decay of priming below in General Discussion.

General Discussion

To sum up, in two simulations, we evaluate HBM against the materials from Pickering and Branigan (1998)'s priming experiment in English ditransitives. The model successfully captures lexical boost, the inverse frequency effect, and asymmetrical decay.

How abstract is syntactic priming?

One of the major questions under debate is: How abstract is syntactic priming? Previous studies observe that priming can be somewhat context-independent, for example, without the alignment of semantic features (Bock & Loebell, 1990; Bock, Loebell, & Morey, 1992) or the specific phonetic content of the morphosyntactic markers (Bock, 1989; Tree & Meijer, 1999). However, in terms of whether priming is verb-independent, the empirical evidence presents a mixed picture. Although there is some evidence for verb-independent syntactic priming from the production side, studies focused on comprehension tend to find that verb overlap is required for a

reliable priming effect (Tooley & Traxler, 2010).

In the current study, by positing a verb-specific level of statistics directly matching the structure of verb-specific input data (i.e. `<give:DO>`), our model assumes that the major locus of syntactic priming lies in the verb-specific biases. At least for the scenario of ditransitives in the current study, this means that what has been directly learned from the input data is the verb-specific statistics, and that there is a lexical basis of syntactic priming where it is the syntactic knowledge stored in the specific verb items that is directly primed. The abstract higher-level priming, under the framework of HBM, can be viewed as an emergent effect that is driven by the verb-specific effects. This idea of lexical-driven syntactic priming is in fact similar to the assumption held by the activation-based mechanism proposed in Pickering and Branigan (1998), where the activation of verb lemmas is first boosted by the prime data, and the activation then spreads to the combinatorial nodes that represent syntactic structures.

Given the assumption that syntactic priming is lexical-driven, *lexical boost* can be reinterpreted as *lexical transfer* by reconsidering the role of verb overlap. If priming is assumed to be operated at an abstract level, it would be reasonable to interpret the role of verb overlap as providing additional processing cues to facilitate the selection of an abstract syntactic structure (Reitter et al., 2011), resulting in a lexical boost. However, if priming is lexical-driven, then the effect of lexical boost, as another side of the same coin, can be interpreted as a weakening effect of verb misalignment. This reinterpretation naturally fits the mechanism of knowledge generalization in HBM: The verb-specific input data not only impacts the verb-specific knowledge, but also goes bottom-up to impact the abstract verb-independent knowledge, which in turn influences other verb-specific knowledge. The magnitude of this generalized priming should be much smaller, not only because the bottom-up effect on the higher-level statistics is counterbalanced by other lower-level statistics, but also because other verb-specific statistics have their own biases.

Decay as interference-based unlearning

What is the mechanism for the decay of priming? In the activation-based account, decay is temporal and in a radioactive fashion. In the implicit learning account, decay is simply underspecified, with no mechanism to unlearn the prime data without any additional input. As intuitive as the time-related decay seems to be, it is empirically difficult to tease apart from another mechanism of decay, which is the interference-based decay. In fact, although the learning literature does not pay much attention to the interference-based decay, studies of working memory has been gradually converged on the claim that memory interference plays a much more important role than temporal decay for many memory mechanisms, such as memory retrieval (Lewis & Vasishth, 2005; Vasishth, Nicenboim, Engelmann, & Burchert, 2019).

The interference-based mechanism in fact can be a potential way to "unlearn" the primed knowledge for implicit learning. In line with the idea that statistical learning spans over life time (Chang et al., 2006), the cognitive system keeps self-updating for the incoming data received after the primes. That being said, for participants in a priming experiment, learning is still ongoing during filler items (Branigan et al., 1999) or in the naturalistic communication environment between experiment sessions (Kaschak, Kutta, & Schatschneider, 2011). This post-priming learning can wash out the original priming effect, manifested as a decay over time.

This interference-based unlearning naturally provides a mechanism for asymmetrical decay in HBM. The lower-level statistics, due to the sparsity of data, can be fairly susceptible to a small amount of incoming data. The higher-level statistics, in contrast, is much more stable as it collects information from all the lower-level statistics. Since the verb-independent priming results from the top-down processing governed by the higher-level statistics, its effect is more stable and less susceptible to the interference from the post-priming data. In another word, although the abstract verb-independent knowledge is more difficult to prime, once it has indeed been primed, it is also more difficult to unprime.

Although we show that HBM can demonstrate the decay of priming in principle, there are still some aspects of decay that can be hardly explained purely from the perspective of implicit learning as a computational theory. First, although HBM captures the qualitative pattern that lexical boost decays faster than verb-independent priming, many studies find that lexical alignment can completely lose its benefit after decay (Hartsuiker et al., 2008). This pattern cannot be reproduced by HBM, where there is always a stronger priming effect with verb overlap. Second, sampling from prior data in the post-priming stage also has its limitation, especially given that the intervening filler items in the experiments of some studies are in a construction differing from the one represented by the statistics in our model. Model architectures that explicitly include a network of constructions as part of the structure of syntactic knowledge, in which (for example) transitive sentences have some relation to ditransitives, may account for the observed decay of syntactic priming when post-priming trials involve sentences in apparently unrelated constructions.

Conclusion

In the current study, we propose a hierarchical Bayesian model for syntactic priming. The model represents syntactic knowledge as hierarchical syntactic statistics, with two levels of abstraction to represent the verb-specific biases and the abstract verb-independent bias for syntactic decisions. As a learning model, HBM is able to capture the three properties of syntactic priming: the lexical boost, the inverse frequency effect, and the asymmetrical decay. This result expands the explanatory power of the implicit learning account to cover more empirical properties of syntactic priming that tend to be accounted for through residual activation, such as the lexical boost and at least some aspects of the short-term priming.

References

Bernolet, S., & Hartsuiker, R. J. (2010). Does verb bias

modulate syntactic priming? *Cognition*, *114*(3), 455–461.

Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, *18*(3), 355–387.

Bock, K. (1989). Closed-class immanence in sentence production. *Cognition*, *31*(2), 163–186.

Bock, K., Dell, G. S., Chang, F., & Onishi, K. H. (2007). Persistent structural priming from language comprehension to language production. *Cognition*, *104*(3), 437–458.

Bock, K., & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, *129*(2), 177.

Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition*, *35*(1), 1–39.

Bock, K., Loebell, H., & Morey, R. (1992). From conceptual roles to structural relations: bridging the syntactic cleft. *Psychological Review*, *99*(1), 150.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (1999). Syntactic priming in written production: Evidence for rapid decay. *Psychonomic Bulletin & Review*, *6*, 635–640.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*(2), B13–B25.

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

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*(7), 294–300.

Ferreira, V. S. (2003). The persistence of optional complementizer production: Why saying "that" is not saying "that" at all. *Journal of Memory and Language*, *48*(2), 379–398.

Fine, A., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS One*, *8*(10), e77661.

Gershman, S. J. (2015). A unifying probabilistic view of associative learning. *PLoS Computational Biology*, *11*(11), e1004567.

Goodman, N. D., & Stuhlmüller, A. (2014). *The Design and Implementation of Probabilistic Programming Languages.* http://dippl.org. (Accessed: 2024-5-3)

Gries, S. T. (2005). Syntactic priming: A corpus-based approach. *Journal of Psycholinguistic Research*, *34*, 365–399.

Hartsuiker, R. J., Bernolet, S., Schoonbaert, S., Speybroeck, S., & Vanderelst, D. (2008). Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language*, *58*(2), 214–238.

Hartsuiker, R. J., & Kolk, H. H. (1998). Syntactic persistence in dutch. *Language and Speech*, *41*(2), 143–184.

Hawkins, R. D., Franke, M., Frank, M. C., Goldberg, A. E., Smith, K., Griffiths, T. L., & Goodman, N. D. (2023). From partners to populations: A hierarchical bayesian account of coordination and convention. *Psychological Review*, *130*(4), 977.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, *127*(1), 57–83.

Kaschak, M. P., Kutta, T. J., & Jones, J. L. (2011). Structural priming as implicit learning: Cumulative priming effects and individual differences. *Psychonomic Bulletin & Review*, *18*, 1133–1139.

Kaschak, M. P., Kutta, T. J., & Schatschneider, C. (2011). Long-term cumulative structural priming persists for (at least) one week. *Memory & Cognition*, *39*, 381–388.

Kemp, C., Perfors, A., & Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical bayesian models. *Developmental Science*, *10*(3), 307–321.

Kleinschmidt, D. F., Fine, A. B., & Jaeger, T. F. (2012). A belief-updating model of adaptation and cue combination in syntactic comprehension. In *Proceedings of the Annual meeting of the Cognitive Science Society* (Vol. 34).

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148.

Levelt, W. J., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psychology*, *14*(1), 78–106.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, *29*, 375–419.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676.

Mahowald, K., James, A., Futrell, R., & Gibson, E. (2016). A meta-analysis of syntactic priming in language production. *Journal of Memory and Language*, *91*, 5–27.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*(3), 283–312.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*(4), 633–651.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin*, *134*(3), 427.

Reitter, D., Keller, F., & Moore, J. D. (2011). A computational cognitive model of syntactic priming. *Cognitive Science*, *35*(4), 587–637.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*(1-3), 107–142.

Scheepers, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition*, *89*(3), 179–205.

Spivey-Knowlton, M., & Sedivy, J. C. (1995). Resolving attachment ambiguities with multiple constraints. *Cognition*,

55(3), 227–267.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285.

Tooley, K. M., & Traxler, M. J. (2010). Syntactic priming effects in comprehension: A critical review. *Language and Linguistics Compass*, *4*(10), 925–937.

Tree, J. E. F., & Meijer, P. J. (1999). Building syntactic structure in speaking. *Journal of Psycholinguistic Research*, *28*, 71–90.

Vasishth, S., Nicenboim, B., Engelmann, F., & Burchert, F. (2019). Computational models of retrieval processes in sentence processing. *Trends in Cognitive Sciences*, *23*(11), 968–982.

Yi, E., Koenig, J.-P., & Roland, D. (2019). Semantic similarity to high-frequency verbs affects syntactic frame selection. *Cognitive Linguistics*, *30*(3), 601–628.

Zhou, Z., & Frank, R. (2023). What affects priming strength? simulating structural priming effect with pips. *Proceedings of the Society for Computation in Linguistics*, *6*(1), 413–417.