A hierarchical Bayesian model for syntactic priming

Weijie Xu & Richard Futrell (UC Irvine)

Syntactic priming refers to the phenomenon that language users tend to repeat to the syntactic decisions they recently made or perceived. Two major accounts are proposed in the literature in order to explain the underlying mechanism of syntactic priming: residual activation[1] and implicit learning[2-3]. The former one well-explains the lexical boost effect of syntactic priming, while the error-based learning mechanism entertained by the latter correctly predicts the inverse frequency effect. The current work aims to show how these different empirical phenomena can be reconciled under a general learning framework, which is the Hierarchical Bayesian Model (HBM)[4-5].

Representing syntactic knowledge in HBM. First, following previous work in syntactic priming / adaptation, we represent syntactic knowledge as syntactic statistics[3,6,7]. Consider the production of English ditransitives. The decision to choose a ditransitive syntactic frame for a given verb can be captured by a Bernoulli distribution, with the parameter representing the bias of Double Object (DO) over Prepositional Object (PO). Second, we propose a hierarchical representation of these syntactic statistics (Fig. 1). The upper level represents the abstract knowledge about the general bias towards DO or PO. The lower level represents the variability of this general bias across different verbs. Therefore, each individual verb imposes its own verb-specific bias ϕ_v.

Learning in the HBM is a Bayesian belief update model as in previous work[6-7]. A common non-hierarchical Bayesian model of adaptation is a Binomial-Beta process where there is a single decision bias Θ whose prior $p(\Theta)$ is a Beta distribution (Eq. 2), and the data likelihood $p(x \mid \Theta)$ follows a Binomial distribution (Eq. 3). In contrast, in our HBM, each individual verb v is associated with a decision bias ϕ_v which is in turn sampled from a general abstract decision bias Θ. Given observed data x_v for verb v, a learner forms the posterior $p(\phi_v, \Theta \mid x_v)$ following (Eq. 4). This verb-specific input data not only impacts the verb-specific parameter ϕ_v, but the effect also goes bottom-up to impact the abstract Θ, which in turn influences the other verb-specific parameters.

Simulation. We evaluate the HBM using the materials from Expt 1 in [1], aiming to capture both the lexical boost and the inverse frequency effect. 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. Each item has two conditions: one has verb overlap between the prime and the target (`Same`); the other does not (`Different`). There are 9 verbs used in their stimuli. We first estimate the pre-priming prior DO probability for each verb based on their frequency and relative use of DO/PO ditransitives in the British National Corpus[8-9]. Then, for each item, the model infer the posteriors given the verb and the corresponding syntactic frame in the prime (Eq. 4). In the end, DO probability for the verb in the target is obtained by marginalizing over the posterior as in (Eq. 9).

Results and discussion. Fig. 2A shows the prior DO probability for the target verbs before priming (`Control`) and the posterior distribution under the two conditions of verb overlap (`Different` vs. `Same`). In general, there is a bias against DO in English ditransitives in priors (DO probability less than 0.5). Fig. 2B shows the priming effect size by comparing the posterior DO probability of the target verbs and their prior probability under the `Control` condition. First, the hierarchical representation of syntactic knowledge helps HBM to successfully capture the lexical boost effect, where the verb overlap induces stronger priming effect. Second, as a typical behavior of Bayesian learning models, HBM captures the inverse frequency effect, where the primes with DO, the originally dispreferred structure, induce stronger effects than PO primes. These results provide a preliminary proof of concept for HBM as a unified cognitive modeling framework for syntactic priming.

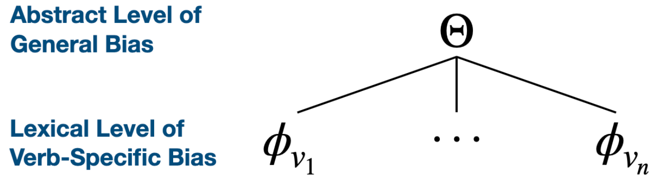

Figure 1: Hierarchical Representation of Syntactic Statistics

Beta-Binomial Process

$$p(\Theta \mid x) \propto p(x \mid \Theta)p(\Theta) \quad (1)$$
$$\Theta \sim \text{Beta}(\alpha, \alpha) \quad (2)$$
$$x \sim \text{Binomial}(\Theta) \quad (3)$$

Hierarchical Bayesian Model

$$p(\phi_v, \Theta|x_v) \propto p(x_v|\phi_v, \Theta)p(\phi_v, \Theta) \quad (4)$$
$$= p(x_v|\phi_v)p(\phi_v|\Theta)p(\Theta) \quad (5)$$
$$\Theta \sim \text{Beta}(1, 1) \quad (6)$$
$$\phi_v \sim \text{Beta}(\alpha\Theta, \alpha(1-\Theta)) \quad (7)$$
$$x_v \sim \text{Binomial}(\phi_v) \quad (8)$$
$$p(\phi_v|x_v) = \int_\Theta p(\phi_v, \Theta|x_v)d\Theta \quad (9)$$

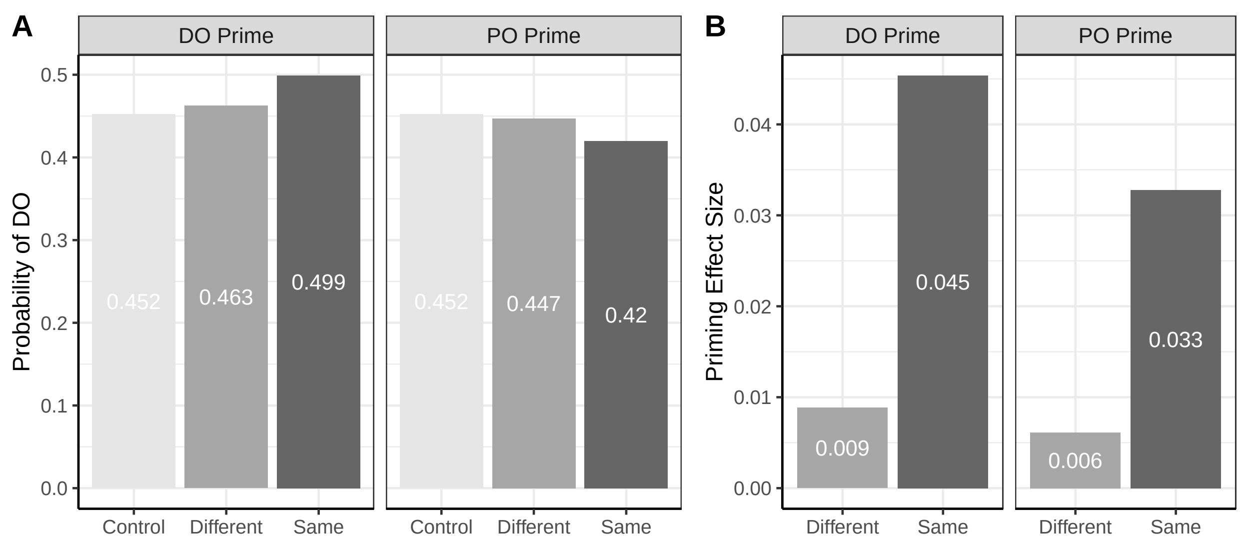

Figure 2: Simulation results on the experiment materials of Pickering & Branigan (1998). *Panel A*: Mean posterior probability of DO for the verbs in target sentences. *Panel B*: Model-predicted priming effect size; y-axis is the difference between the posterior and the prior probability of DO for the verbs in target sentences.

References. [1] Pickering & Branigan (1998) *JML*; [2] Chang et al. (2006) *Psych Review*; [3] Jaeger & Snider (2013) *Cognition*; [4] Tenenbaum et al. (2015) *Science*; [5] Kemp et al. (2007) *Dev Sci*; [6] Fine et al. (2013) *PloS One*; [7] Kleinschmidt et al. (2012) *CogSci*; [8] Yi et al. (2019) *Cog Ling*; [9] Zhou & Frank (2023) *SCiL*