Korean Journal of English Language and Linguistics, **Vol 25, March 2025, pp. 415-430**
DOI: 10.15738/kjell.25..202503.415

KOREAN JOURNAL OF ENGLISH LANGUAGE AND LINGUISTICS

ISSN: 1598-1398 / e-ISSN 2586-7474

http://journal.kasell.or.kr

On the Properties of Syntactic Priming: A L2 Hierarchical Bayesian Model Approach*

Sunjoo Choi · **Myung-Kwan Park** (Dongguk University)

This is an open-access article distributed under the terms of the Creative Commons License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received: February 15, 2025
Revised: March 7, 2025
Accepted: March 17, 2025

Choi, Sunjoo (first author)
Research Professor, Division of English Language and Literature, Dongguk University
Email: sunjoo@dongguk.edu

Park, Myung-Kwan
(corresponding author)
Professor, Division of English Language and Literature, Dongguk University
Email:parkmk@dongguk.edu

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea.
(NRF-2023S1A5B5A16080469)

ABSTRACT

Choi, Sunjoo and Myung-Kwan Park. 2025. On the properties of syntactic priming: A L2 hierarchical Bayesian model approach. *Korean Journal of English Language and Linguistics* 25, 415-430.

A range of empirical factors have been identified in the literature as interacting with the strength of syntactic priming: the lexical boost, the inverse frequency effect, and the asymmetrical decay. This study explores how these factors can be represented within a general learning framework called the hierarchical Bayesian model (HBM), utilizing data from the K-English Textbook corpus. The HBM conceptualizes syntactic knowledge as a hierarchical structure of syntactic statistics, which is continually updated through Bayesian inference based on the language experience (Xu and Futrell 2024). Given this background, the current research aims to investigate the underlying mechanism of syntactic priming from a different angle using statistical learning. After building the L2 HBM, two simulations are conducted employing Pickering and Branigan's (1998) English ditransitive materials. In so doing, we demonstrate that the L2 HBM successfully captures the aforementioned properties of syntactic priming, as a previous study reported. To account for these observed factors simultaneously, we support the claim that empirical properties of syntactic priming are realized in the cognitive model architecture.

KEYWORDS

syntactic priming, lexical boost, inverse frequency, asymmetrical decay, hierarchical Bayesian model, syntactic statistics, probability, ditransitives

© 2025 KASELL All rights reserved

1. Introduction

Syntactic priming is a widely used paradigm in psycholinguistics to examine an adaptive behavior where exposure to a specific syntactic structure influences a language user's subsequent processing of congruent structures. Simply speaking, speakers tend to reproduce similar syntactic structures they have recently encountered (Bock 1986, Bock and Loebell 1990, Branigan et al. 2000, Pickering and Ferreira 2008). This paradigm allows us to investigate the extent to which the production or processing of target sentences is facilitated by preceding prime sentences. The following example illustrates a target sentence like (1).

(1) The girl will send the boy the letter.

Target sentences such as (1) are produced more often or are processed more quickly when they are preceded by prime sentences such as (2), which have a congruent structure, than when they are preceded by the prime sentence such as (3). Example (3) describes the same transfer event as in the example of (2) but has an incongruent structure with the sentence (2).

(2) The boy will give the girl the ball.
(3) The boy will give the ball to the girl.

A speaker is more likely to produce a double-object (DO) construction (1) after being exposed to a sentence that has a congruent structure (2) rather than after encountering a sentence with an incongruent structure (3). Simply speaking, the probability of using a double-object (DO) construction is higher than that of using a prepositional object (PO) construction. There is a substantial increase in the usage of the same syntactic form to describe a subsequent ditransitive event. The effect of syntactic priming is well-attested in humans for language production (Mahowald et al. 2016) and comprehension (Tooley 2023). Recently, significant priming effects have been observed in large language models (Choi and Park 2022, Jumelet et al. 2024, Michaelov et al. 2023, Prasad et al. 2019, Sinclair et al. 2022, van Schijndel and Linzen 2018). Large language models have exhibited human-like manner in terms of syntactic priming.

Significant priming effects are generally observed when there are shared words between prime and target sentences. Also, the effects are associated when the prime sentence is more unusual or less frequent. Based on these findings, several well-established factors that enhance priming effects. (Mahowald et al. 2016). These include: (i) Lexical Boost: When the prime and target sentences share key lexical items, the priming effect is significantly increased; (ii) Inverse Frequency: Less frequent syntactic constructions tend to produce stronger priming effects; (iii) Asymmetrical Decay: lexical boost decays faster than the abstract verb-independent priming effect. This paper concentrates on these three empirical properties and the probabilistic nature of priming. In what follows, we will specifically explain these three factors.

The lexical boost describes the phenomenon where the size of the syntactic priming effect increases when the prime sentence and the target sentence include similar lexical items (Pickering and Branigan 1998). For instance, if both the prime and target sentences utilize the same verb, such as "*send*," the priming effect for the DO construction is significantly stronger compared to a target sentence where the prime and target sentences employ different verbs, like "*loan*."

Another significant characteristic of syntactic priming is the inverse frequency effect, which suggests that constructions that are less frequently encountered in language tend to produce a stronger priming effect (Bock

© 2025 KASELL All rights reserved

1986, Ferreira 2003, Scheepers 2003, Hartsuiker and Kolk 1998, Kaschak et al. 2011). For example, between the double-object (DO) and prepositional-object (PO) constructions, PO constructions are less common. As a result, a PO prime significantly enhances the likelihood of a PO target being used compared to the effect of a DO prime on a DO target. Jaeger and Snider (2013) found that the magnitude of the priming effect is directly linked to the prediction error experienced during processing, which is affected by the statistical patterns evident in language use. Therefore, constructions that are less frequent or co-occur less often generate larger prediction errors, resulting in a stronger priming effect.

The final phenomenon of interest relates to the reduction in the size of the priming effect over time. The decay in syntactic priming has been observed in both short-term (Branigan et al. 1999, Levelt and Ketler 1982) and long-term experiments (Bock et al. 2007, Bock and Griffin 2000, Kaschak et al. 2011). Specifically, asymmetrical decay refers to the finding that the lexical boost decreases more quickly than the abstract priming effect when there is no lexical alignment between the prime and the target (Bock and Griffin 2000, Hartsuiker et al. 2008).

In general, the syntactic priming paradigm entails a decision-making process in which individuals choose between various syntactic structures to express the same message during production or interpret multiple potential parses when encountering relevant input. In this sense, the decision-making processing can be generalized using a probability distribution over the relevant syntactic decisions. For instance, in the case of ditransitives constructions, the decision between the DO form and the PO form can be seen using a Bernoulli distribution, in which the parameter indicates the likelihood of producing the DO form instead of the form PO. This perspective is the reason why we adopt statistical methodology for the research. In this vein, representing a person's decision processing is another way to investigate the syntactic priming paradigm. In other words, counting how often one expects to experience different syntactic congruent constructions is required. In doing so, we can demonstrate this probability distribution and understand how it changes through experience or learning.

In this paper, we seek to better understand the mechanism that may underlie syntactic priming behavior. To probe the characteristics of syntactic priming, we assume that it is essential to conceptualize an individual's syntactic knowledge as syntactic statistics. For this reason, we adopt Xu and Futrell's computational-level theory of syntactic priming based on a hierarchical Bayesian model (HBM) of syntactic knowledge. Borrowing insights from HBM, we explore how and why higher probabilities are assigned to target sentences depending on input data. Our L2 HBM builds upon the fundamental Bayesian belief-update model by representing syntactic knowledge across multiple levels of abstraction using the K-English Textbook corpus[1] (Kemp et al. 2007). As a general learning framework, hierarchical Bayesian models have been widely employed in various linguistic fields, including speech perception (Kleinshmidit and Jaeger 2015), the development of communicative systems (Hawkins et al. 2023), and human cognition (Tenenbaum et al. 2011). Given its verified performances, we show how this general learning framework employing the K-English Textbook corpus can capture the three empirical factors of syntactic priming: the lexical boost effect, the inverse frequency effect, and asymmetrical decay. Then, we compare the results with those of their previous counterparts and suggest an account for them simultaneously.

The paper is organized as follows. Section 2 reviews Xu and Futrell's (2024) recent study of the issue in focus. Especially, the Beta-Binomial process is introduced in greater detail. Section 3 adds to the L2 hierarchical Bayesian model and reports the results of two simulations. Section 4 discusses the observed behaviors from a series of simulations, and two relevant well-known theories regarding the empirical factors of priming are noted. Section 5

[1] K-English Textbook corpus: Korean L2 learning materials, specifically collected from EBS-CSAT English Prep books (2016-2018) and the English textbooks used in middle and high schools (2001-2002, 2009-2010, 2015-16) and comprises approximately 3.1 million tokens (Choi 2024, Choi and Park 2022).

© 2025 KASELL All rights reserved

wraps up with a conclusion.

2. Xu and Futrell (2024)

Recently, Xu and Futrell (2024) conceptualize a person's syntactic knowledge using syntactic statistics, examining the frequency with which different syntactic constructions are anticipated at various levels of abstraction. As depicted in Figure 1, they propose a hierarchical structure for these syntactic statistics, consisting of two levels of abstraction. The lower level means verb-specific statistics ϕ_v, each verb imposes its own bias (Bernolet and Hartsuiker 2010). The verb-specific statistics echo the lexicalist view of sentence processing, where syntactic knowledge is accessible via lexical representations (MacDonald et al. 1994, McRae et al. 1998, Spivey-Knowlton and Sedivy 1995). Conversely, the higher-level abstraction, Θ, aggregates all the lower-level statistics and reflects a general decision bias applicable to all verbs.

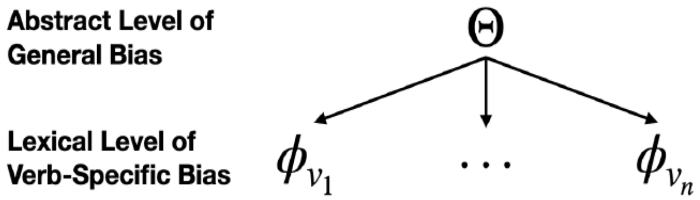

Figure 1. Hierarchical Representation of Syntactic Statistics (Xu and Futrell 2014, p. 3)

For instance, the variables illustrated 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). In the context of ditransitive priming, these variables would reflect the frequency of occurrences for DO and PO constructions, based on either individual verbs or an overall aggregation across multiple verbs. This structure of syntactic statistics is not restricted to ditransitive constructions. The mechanism can be utilized for any constructions that involve processing decisions, such as garden-path sentences that require ambiguity resolution. It can also apply to the decision of syntactic frames in production, including choices between passive and active structures in transitive constructions. Xu and Futrell suggest that syntactic learning occurs through Bayesian updates within the hierarchical model presented in Figure 1. To elucidate how this mechanism functions, we first explain its application using a well-known non-hierarchical Beta Binomial model and subsequently explore the complete hierarchical model.

First, the Beta-Binomial model serves as a fundamental example of Bayesian belief updating. It establishes a generative model in which a set of decisions *x* regarding two possible outcomes is drawn from a Binomial distribution, which is parameterized by the variable Θ. This parameter Θ is then sampled from a prior Beta distribution defined by hyperparameters α and β, with both α and β greater than zero (α, β > 0). This framework enables the integration of prior knowledge with observed data to update beliefs about the likelihood of the outcomes.

(4) Θ ~ Beta (α,β)

(5) x ~ Binomial (Θ)

© 2025 KASELL All rights reserved

In the realm of syntactic priming, the variable Θ denotes the probability of the DO form in comparison to the PO form, and the hyperparameters α, β in the Beta distribution reflect prior expectations regarding the frequencies of the DO and PO outcome respectively. This framework enables the updating of these probabilities in response to observed production data, thereby enhancing the understanding of patterns in syntactic choice.

(6) $p\ (\Theta \mid \chi) \propto p\ (\chi \mid \Theta)\ p\ (\Theta)$.

This updated distribution follows a straightforward form. If one observes N datapoints containing x examples of a DO outcome, the updated decision probabilities are outlined as follows:

(7) $\chi \mid \Theta \sim \text{Beta}\ (\alpha + \chi,\ \beta + N - \chi)$.

As a result, the observation of numerous DO outcomes will enhance the likelihood of future DO probabilities. In this framework, Xu and Futrell propose a hierarchical structure for syntactic statistics. They argue that each verb possesses an individual decision bias, denoted as ϕ_v, and they introduce a generative model to illustrate this concept.

(8) $\Theta \sim \text{Beta}\ (1, 1)$
(9) $\phi_v \sim \text{Beta}\ (\alpha\Theta,\ \alpha(1- \Theta))$
(10) $\chi_v \sim \text{Binomial}\ (\phi_v)$

In a generative model, a global decision bias variable Θ is drawn from a distribution Beta (1,1). Subsequently, each verb *v*, an individual decision bias ϕ_v is drawn from a Beta distribution that is influenced by the variable Θ. In this vein, the parameter α, which is greater than 0, indicates the significance of the global prior Θ in determining the verb-specific biases ϕ_v. The hierarchical framework of the generative model suggests that interacting with one verb can lead to changes not just in that specific verb's decision bias, but also in the global decision bias, which subsequently influences other verbs. Based on the counts of χ_v DO outcomes for a particular verb v when focusing on updating the statistics for the currently observed verb *v*, this update is accomplished by applying Bayes' rule only to ϕ_v. This is described in (11):

(11) $p(\phi_v \mid \chi_v) \propto p\ (\chi_v \mid \phi_v)\ p\ (\phi_v)$

When updating the variable ϕ_w for a different verb, then the update is facilitated by the revised global statistics Θ:

(12) $p(\phi_w \mid \chi_v) = \int p(\phi_w \mid \Theta)\ p\ (\Theta \mid \chi_v)\ d\Theta$[2]

As a result, information in the HBM flows both bottom-up and top-down. When new data χ_v for a verb is encountered, it affects the verb-specific parameter ϕ_v and the effect goes bottom up to impact the higher-level Θ.

[2] The formulas (4~12) are illustrated in Xu and Futrell (2024, p. 3).

© 2025 KASELL All rights reserved

Since this abstract Θ governs and constrains all the verb-specific parameters as a top-down process, the effect on Θ in turn influences other ϕ_w. Thus, learning based on the data associated with one verb can be applied to others. To facilitate this learning process within the HBM, they employ WEBPPL[3], as noted by Goodman and Stuhlmüller (2014).

Xu and Futrell conducted an evaluation of their model using the experimental materials from Experiment 1 by Pickering and Branigan (1998) through two simulations. The purpose of the first simulation is to capture both the lexical boost effect and the inverse frequency effect. To achieve this, they construct a dataset that reflects the language experiences of participants prior to the priming experiment. As shown in Table 1, this dataset includes 100 data points distributed among the nine verbs featured in the original experiment. Each data point shows an instance of a verb used in one of the two ditransitive structures (e.g., <send: DO>, <lend: PO>). The frequency of each verb and their relative use of DO/PO are specified based on counts from the British National Corpus (Yi et al. 2019, Zhou and Frank 2023). The model starts with the global bias Θ and verb-specific biases ϕ_v derived from this dataset in accordance with the previously described procedures. This initial state is the prior distribution, which will be compared with the updated posterior distribution after exposure to the priming data.

Table 1. Verb Frequencies and Counts of DO/PO Ditransitives Used to Form the Model's Prior Distribution (Xu and Futrell 2024, p. 4)

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

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 assessed from the behavioral data are the verb-specific parameters, which correspond to the proportion of DO/PO usage associated with the specific verbs in the target sentences of each item. Consequently, for each target verb, the verb specific posterior is obtained through marginalization, as indicated in equation (12).

[3] http://webppl.org

© 2025 KASELL All rights reserved

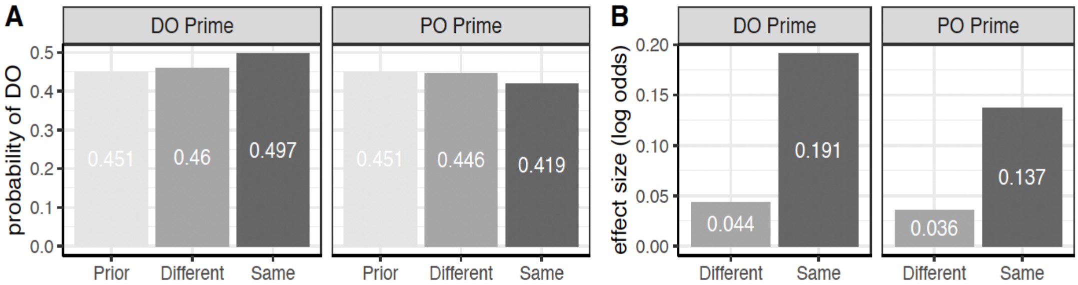

Figure 2: Results of Simulation 1 (Xu and Futrell 2024, p. 4)[4]

Figure 2A shows the average DO probability for the target verbs across various conditions. The control condition illustrates the prior DO probability prior to priming. Despite the raw frequency data in Table 1 indicating a higher number of DOs, the model identifies a global bias that slightly against DO (i.e. $P(\text{DO}) < 0.5$). This is primarily due to the fact that the DO preference in the raw frequency counts is heavily influenced by one dominant verb, "*give*," within the prior dataset. The model demonstrates a priming effect, characterized by an increased probability of DO usage when exposed to DO primes and a decreased probability with PO primes.

Figure 2B provides the priming effect size more directly by calculating difference of log-odds between the prior and the posterior of DO. Additionally, the model predicts a stronger priming effect when the prime and the target verb are identical. Interestingly, it also demonstrates a significant priming effect from DO primes, even though these primes were less favored in the prior data.

The objective of the second simulation is to investigate the model of the asymmetrical decay of syntactic priming. The procedure followed in this simulation is consistent with the methodology used in the first simulation, using the data provided in Table 1. The simulation is divided into two groups: an exposure group and a control group. In the exposure group, the model is initially conditioned on the prime sentence, whereas the control group does not receive any conditioning on prime sentences. After this initial phase, both groups receive two batches of additional data, each consisting of 100 samples drawn from the frequency distribution outlined in Table 1. These post-priming data represent the average effects of post-priming trials and experiences. The size of the priming effect is then determined by calculating the difference in log-odds for DO between the model-inferred posterior in the exposure group and that in the control group. Figure 3 illustrates the model-predicted size of the priming effect in relation to the number of batches of post-priming data[5].

[4] *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.
[5] The result with zero batches of post-priming data reflects the finding from Simulation 1 (Figure 2B) and is included here for convenient comparison.

© 2025 KASELL All rights reserved

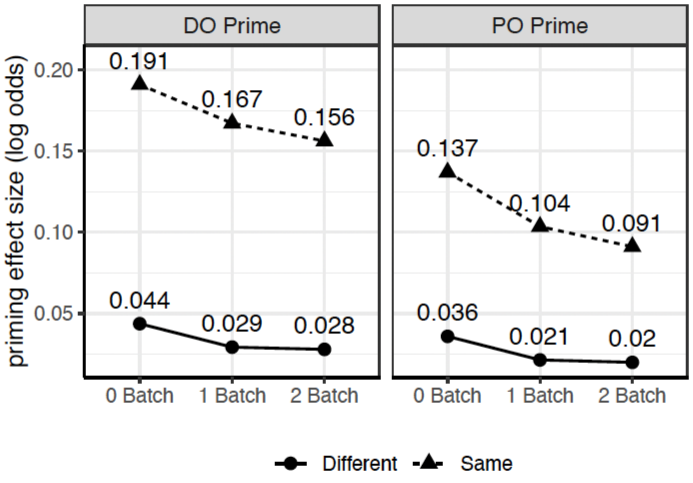

Figure 3. Simulation 2 Model-Predicted Priming Effect Size (Xu and Futrell 2024, p. 5)[6]

Firstly, consistent with the first simulation, the model effectively demonstrates both the lexical boost and the inverse frequency effect when additional post-priming data is introduced. Secondly, the results demonstrate a general decline in the priming effect across both conditions of verb overlap, indicating that the size of the priming effect diminishes as the model incorporates more post-priming data. The size of the priming effect decreases more than twice as quickly in the condition with verb overlap compared to the condition without verb overlap.

Conducting two simulations, Xu and Futrell assess the HBM using the materials from Pickering and Branigan's (1998) priming experiment involving English ditransitives. The model effectively captures the lexical boost, the inverse frequency effect, and the asymmetrical decay. Based on these observed behaviors, we further investigate the issues by training the model based on the K-English Textbook corpus. In what follows, we will introduce how we reorganize the methodology and then report the behaviors conducted by a series of simulations.

3. The L2 Hierarchical Bayesian Model and its Ability

A critical issue currently under discussion is the level of abstraction regarding syntactic priming. Previous studies have shown that syntactic priming can demonstrate a certain degree of context independence, occurring without the need for alignment of semantic features (Bock and Loebell 1990, Bock et al. 1992) or specific phonetic content of morphosyntactic markers (Bock 1989, Tree and Meijer 1999). However, the empirical evidence regarding the verb independence of priming is controversial. While some studies support the idea of verb-independent syntactic priming from a production perspective, other research often suggests that verb overlap is crucial for achieving a consistent priming effect (Tooley and Traxler 2010).

In light of this issue, this research adopts a computational framework for understanding syntactic priming using a hierarchical Bayesian model (HBM) of syntactic knowledge. As illustrated before, this statistical model is built upon the standard Bayesian belief-update approach by incorporating multiple levels of abstraction (Kemp et al.

[6] It is 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.

© 2025 KASELL All rights reserved

2007). This study illustrates how this framework updating with the K-English Textbook corpus effectively captures three empirical factors of syntactic priming[7]. The work aims to compare this probability distribution with a previous study and investigate how it changes based on experiential factors. Building on prior research, we conceptualize an individual's syntactic knowledge as syntactic statistics: counts reflecting the expected occurrences of different syntactic constructions, which can be represented at varying levels of abstraction. As illustrated in Figure 1, two distinct levels of abstraction within this framework are suggested.

The first step of experiment is to construct a dataset that reflects the persons's linguistic experience prior to the priming experiment. As detailed in Table 2, this dataset consists of 100 data points distributed across the nine verbs used in the original study. The frequency of each verb, along with its relative usage of DO and PO constructions, is based on counts derived from the K-English Textbook corpus. The verb frequencies in Table 2 are drawn from a multinomial distribution with the parameter $N = 100$ and the parameter 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. This initial state represents the prior distribution, which will be compared to the updated posterior distribution after exposure to the priming data.

Table 2. Verb Frequencies and Counts of DO/PO Ditransitives Used Form the L2 model's Prior Distribution

Verb	DO Freq.	PO Freq.	Verb Freq
give	41	22	63
show	4	8	12
send	2	9	11
lend	3	2	5
hand	1	1	2
loan	0	2	2
offer	0	3	3
sell	0	2	2
post	0	0	0
in total	51	49	100

In two simulations, we also assess the model using the experimental materials from Pickering and Branigan's (1998) Experiment 1, aiming to capture the three empirical properties of syntactic priming mentioned earlier. The original experiment employs a trial-to-trial production priming paradigm involving English ditransitives[8]. There are 32 items, each comprising a prime and a target, as illustrated in (11) below. The prime is biased towards either a direct object (DO) structure or a prepositional object (PO) structure. Furthermore, verb overlap is adjusted, so that the prime either shares the same verb as the target or a different verb.

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

a. Prime:

The captain gave the spare lifejacket …

The captain gave old sailor …

The captain lent the spare life jacket …

---

[7] See Footnote 1 for the details about the corpus.

[8] At the prime, participants are presented with the partial 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.

© 2025 KASELL All rights reserved

The captain lent the old sailor …
b. Target: The racing driver sent ...

Considering this background, the condition involving verb overlap in Pickering and Branigan will yield a stronger priming effect in the model output compared to the condition without verb overlap. The following sub-section reports the results of the first simulation.

3.1 Simulation 1

Building on the prior, the model infers a posterior distribution using the exposure data presented in the primes from Pickering and Branigan (1998). As mentioned in Section 2, the posterior learned by the L2 model is a joint distribution for both the global and verb-specific decision biases. For the target verb in each item, we derive the verb-specific posterior by applying marginalization, as outlined in Equation 12.

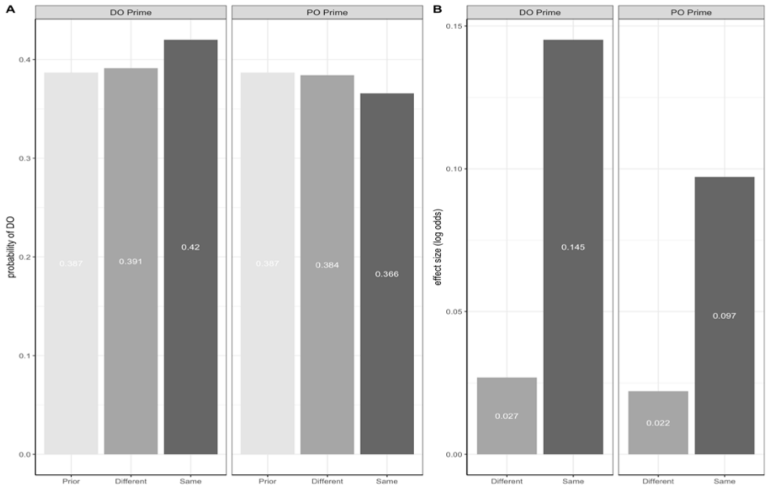

Figure 4. Results of Simulation 1 of the L2 HBM

Figure 4A illustrates the average DO probability for the target verbs under condition. In the control condition, this reflects the prior DO probability before priming. Despite the raw frequencies in Table 2 showing a higher number of DO instances, the model infers a slight global bias against DO (i.e., $p(\text{DO}) < 0.5$)[9]. This pattern is mainly influenced by one single frequent verb. After encountering the prime data, a priming effect is observed. As predicted, we found an increase in DO probability with DO primes and a decrease with PO primes. On the right side of the graph, figure 4B highlights the priming effect size more explicitly by presenting the log-odds difference between the prior and posterior DO probabilities for each target verb.

In line with the lexical boost effect, the model demonstrates a stronger priming effect when the prime and target share the same verb. In other words, if both the prime and the target involve the same verb "sell" then there is an increase in the effect size. In accordance with the inverse frequency effect, the model predicts that DO primes, the less favored structure in the prior data, produced a stronger priming effect. As we assumed, the unexpected

[9] This occurs because the apparent DO preference in the raw data is primarily influenced by a single highly frequent verb, "*give*," in the prior dataset.

© 2025 KASELL All rights reserved

prediction induces more significant effects. The observed patterns are similar to those reported in Xu and Futrell's first simulation.

As mentioned, the model's capacity to account for the inverse frequency effect is based on the principles of Bayesian learning. Also, the L2 HBM shows the lexical boost effect. We can safely conclude that two HBMs prove that the models clearly capture the lexical boost and the inverse frequency effect when encountering additional priming data. It is important to note that the model effectively replicated conclusions drawn by Pickering and Branigan. In what follows, we will examine the last factor that modulates syntactic priming effects: the asymmetrical decay.

3.2 Simulation 2

In Simulation 2, we focus on modeling the asymmetrical decay in syntactic priming, where the lexical boost diminishes more quickly than the abstract priming effect that does not involve verb overlap. The estimation process follows that of Simulation 1, utilizing the prior data listed in Table 2. As the previous study did, the second simulation is divided into two groups: an exposure group and a control group. The model for the exposure group is initially conditioned on the prime sentences, while the control group does not undergo such conditioning. Following this, both groups are presented with two sets of additional data, each consisting of 100 samples drawn from the frequency distribution in Table 2. These post-priming datasets represent the average impact of post-priming trials and experiences. Finally, the size of the priming effect is calculated as the difference in log-odds for DO between the inferred posterior of the exposure group and that of the control group. The results are shown in Figure 5.

Figure 5 represents the size of the priming effect predicted by the model, plotted against the number of batches of post-priming data used. The result with zero batches of post-priming data aligns with the findings (shown in Figure 4B) and is included here for comparison. Initially, similar to the results in Simulation 1 of the L2 HBM, the model accurately reflects the lexical boost and the inverse frequency effect when additional post-priming data is included. Additionally, the findings suggest that there is a general decline in the priming effect across both verb overlap conditions. As the model processes greater amounts of post-priming data, the size of the priming effect decreases.

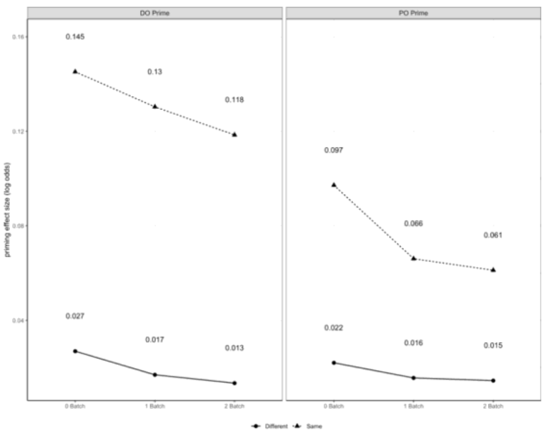

Figure 5. Results of Simulation 2 of the L2 HBM

© 2025 KASELL All rights reserved

It is noteworthy that the model captures an asymmetrical decay. In other words, the decrease in priming effect size occurs at more than double the rate in the condition where verbs overlap compared to the condition where they do not. All things considered, the L2 HBM effectively demonstrates the asymmetrical longevity of syntactic priming, illustrating that the lexical boost effect diminishes more quickly than the priming effect observed in the absence of verb overlap. In the following discussion section, we investigate the mechanism of how HBM can address the asymmetrical decay of priming.

Through two simulations, the L2 HBM's observed behaviors support the previous findings that represented the syntactic knowledge as syntactic statistics. The findings are identical to Xu and Futrell's findings. Following two simulations, we also show how three empirical phenomena can be reconciled under a general learning framework. Our results provide preliminary validation for using Hierarchical Bayesian Models as a coherent cognitive modeling framework for understanding the underlying mechanism of syntactic priming. Also, we agree that a hierarchical representation of syntactic statistics, as represented in Figure 1, is a novel methodology that expands to cover well-documented properties of syntactic priming.

4. Discussion

The currently-debated issue of syntactic priming is which factors strengthen its effects. Earlier studies have noted that the priming effect can be context-independent, occurring even in the absence of shared semantic features (Bock and Loebell 1990, Bock et al. 1992) or specific phonetic elements of morphosyntactic markers (Bock 1989, Tree and Meijer 1999). Nevertheless, the empirical evidence concerning whether the syntactic priming effect is context-independent has been controversial. Although some studies on production provide evidence for verb-independent syntactic priming, research on comprehension typically suggests that verb overlap is required to reach a significant priming effect (Tooley and Traxler 2010).

To elucidate the issue, we have investigated three factors that modulate the strength of syntactic priming. Materials from Pickering and Branigan's (1998) experiment on syntactic priming in English ditransitives were adopted to conduct two simulations. This paper has used a reliable methodology to represent syntactic knowledge within a hierarchical structure. Many previous works based on neural language models (NLMs) noted that NLMs are susceptible to syntactic priming and its effects are strongly boosted by various factors such as semantic similarity and lexical overlap (Mahowald et al. 2016, Sinclair et al. 2021). Given these verified performances, we have looked into boosting factors for the syntactic priming from a different angle. For this specific aim, we have adopted a statistical methodology for the experiment. As described in the previous section, the syntactic priming paradigm involves a cognitive decision-making process where individuals select from different syntactic structures to convey the same message during language production. In this regard, the decision-making process can be illustrated through a probability distribution that reflects the likelihood of making different relevant syntactic choices. For this reason, following Xu and Futrell's statistical approach based on a general learning framework, we have collected a prior distribution drawn from K-English textbook corpus and conducted simulations and then compared it with the updated posterior distribution. As a result, the L2 HBM effectively captured three properties of priming: the lexical boost, the inverse frequency effect, and asymmetrical decay. These observed behaviors were consistent with Xu and Futrell's results. Taken together, these two HBMs have successfully captured well-documented empirical factors.

Based on these HBM behaviors, we have assumed that the crucial factor of syntactic priming stems from verb specific biases by introducing a level of statistics that aligns with the structure of verb specific input data. This

© 2025 KASELL All rights reserved

means that the model learns verb specific statistics from the exposed data, creating a lexical foundation for syntactic priming where the syntactic knowledge related to specific verbs is activated or encountered. Within the HBM framework, higher-level abstract priming can be viewed as an emergent effect influenced by these verb-specific factors. This notion of lexical-driven syntactic priming is relevant to the activation-based mechanism proposed by Pickering and Branigan (1998). The activation of verb lemmas is initially enhanced by prime data, which then spreads to the combinatorial nodes representing syntactic structures.

As noted by Xu and Futrell, assuming syntactic priming is influenced by lexical factors, the lexical boost can be reinterpreted as a form of lexical transfer. If priming functions at an abstract level, then it makes sense to consider verb overlap as providing additional processing cues that facilitate the selection of an abstract syntactic structure leading to a lexical boost (Reitter et al. 2011). On the other hand, if priming is indeed driven by lexical factors, the lexical boost effect might be interpreted as a weakening effect resulting from verb misalignment. This reinterpretation aligns with the knowledge generalization mechanism in HBM, where verb-specific input data not only affects verb-specific knowledge but also goes a bottom-up. Then, it turns to influence verb-specific knowledge.

It is worth noting that two major theoretical approaches have proposed in the literature to explain the underlying mechanism of syntactic priming. The residual activation theory of priming suggests that the production or comprehension of a specific syntactic structure is impacted by its level of activation. When a sentence is processed, there is an increase in activation for both the lexical items and the syntactic structure, which facilitates in their subsequent reuse (Pickering and Branigan 1998). Pickering and Branigan describe a representational structure in which nodes that denote verb lemmas are linked to combinatorial nodes that represent syntactic structures. In this framework, an increase in activation of a particular verb lemma propagates to the combinatorial nodes as well as to other verb lemmas, which subsequently enhances the activation of the combinatorial nodes. This residual activation account accounts for the lexical boost effect, as the verb lemma presented in the prime is directly activated and therefore receives the most significant boost in activation. Unfortunately, the residual activation account struggles to account for long-term priming, as it assumes that all activation diminishes over time.

The implicit learning theory suggests that syntactic priming can be a form of syntactic learning. This idea was initially implemented in a connectionist model that predicts the next word based on what has been previously encountered (Chang et al. 2006). In this framework, the prediction error is a crucial learning signal, which updates the connection weights between nodes through backpropagation. According to the implicit learning theory, the priming effect size is proportional to the prediction error when processing the prime, thereby explaining the inverse frequency effect (Jaeger and Snider 2013). However, in contrast to the residual activation theory, the implicit learning account finds it challenging to explain the observation that the syntactic priming effect can occasionally be temporary.

Unfortunately, neither account can fully explain all three properties simultaneously. Interestingly, Reitter et al. (2011) present a hybrid approach explaining both residual activation and implicit learning. They introduce an ACT-R model, the memory-based model, of syntactic priming based on a wide-coverage, lexicalized syntactic theory that explains priming as facilitation of lexical access. The model presented here specifies how information is encoded in memory, retrieved, and processed. ACRT-R offers two basic mechanisms that can account priming, that is base-level learning and spreading activation. Through a series of simulations, their model embodies the claim that priming applies to syntactic structure, in the form of combinatorial categories as syntactic descriptions of subcategorization properties. In a nutshell, the primary contribution is to show that syntactic priming can be understood as the result of two complementary learning effects: the formation of individual syntactic representations and the development of connections between these syntactic representations and associated lexical

© 2025 KASELL All rights reserved

or semantic material. In terms of the asymmetrical decay, lexical chunks are evacuated from working memory, they cannot spread activation anymore. In the model presented here, syntactic priming is assumed to be a result of more general cognitive phenomena affecting syntactic processing (Pickering and Branigan 1998, Pickering et al. 2002). Even though the model presented here focuses on the aspects of language production involved in syntactic priming, its cognitive architecture can explain all the three empirical properties. The L2 results provide evidence of this hybrid approach in explaining the properties of syntactic priming[10].

5. Conclusion

In this study, we provide novel evidence using a hierarchical Bayesian model (HBM), which is implemented to explain the nature of syntactic priming. Borrowing insights from a statistical approach, the model conceptualizes syntactic knowledge based on hierarchical syntactic statistics featuring two levels of abstraction. From a different angle, a general learning framework is needed to investigate the nature of syntactic priming. We built the L2 HBM based on the K-English Textbook corpus and evaluated the model using English ditransitives through two simulations. Consequently, the L2 HBM captured three critical properties of syntactic priming: the lexical boost, the inverse frequency effect, and the asymmetrical decay. These results support the effectiveness of HBM as an integrated cognitive modeling framework for syntactic priming. To explain the underlying mechanism of syntactic priming simultaneously, we follow the claim that priming effects are driven by two different learning mechanisms based on well-established general cognitive principles: the learning of syntactic representations and their connection with lexical information.

References

Bernolet, S. and R. J. Hartsuiker. 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. and H. Loebell. 1990. Framing sentences. *Cognition* 35(1), 1-39.

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

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

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

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

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

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

Choi, S. J. and M. K. Park. 2022. Syntactic priming in the L2 neural language model. *The Journal of Linguistic*

[10] Also, we leave the comparison experiment between humans and HBM using a general learning framework in future works.

© 2025 KASELL All rights reserved

Science 103, 81-104.
Choi, S. J. 2024. On Relative Clause Attachment Preferences in the L2 LSTM LM. *Studies in Modern Grammar* 122, 213-234.
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.
Goodman, N. D. and A. Stuhlmüller. 2014. The Design and Implementation of Probabilistic Programming Languages. http://dippl.org. (Accessed: 2024-5-3)
Hartsuiker, R. J. and H. H. Kolk. 1998. Syntactic persistence in dutch. *Language and Speech* 41(2), 143-184.
Hartsuiker, R. J., S. Bernolet., S. Schoonbaert., S. Speybroeck and D. Vanderelst. 2008. Syntactic priming persists while the lexical boost decays: Evidence from written and spoken dialogue. *Journal of Memory and Language* 58(2), 214-238.
Hawkins, R. D., M. Franke., M. C. Frank., A. E. Goldberg., K. Smith., T. L. Griffiths and N. D. Goodman. 2023. From partners to populations: A hierarchical bayesian account of coordination and convention. *Psychological Review Review* 130(4), 977.
Jaeger, T. F. and N. E. Snider. 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.
Jumelet, J., W. Zuidema and A. Sinclair. 2024. Do Language Models Exhibit Human-like Structural Priming Effects?. arXiv preprint arXiv:2406.04847.
Kaschak, M. P., T. J. Kutta and J. L. Jones. 2011. Structural priming as implicit learning: Cumulative priming effects and individual differences. *Psychonomic Bulletin & Review* 18, 1133-1139.
Kemp, C., A. Perfors. and J. B. Tenenbaum. 2007. Learning overhypotheses with hierarchical bayesian models. *Developmental Science* 10(3), 307-321.
Kleinschmidt, D. F. and T. F. Jaeger. 2015. Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological review* 122(2), 148.
Levelt, W. J. and S. Kelter. 1982. Surface form and memory in question answering. *Cognitive Psychology* 14(1), 78-106.
MacDonald, M. C., N. J. Pearlmutter and M. S Seidenberg. 1994. The lexical nature of syntactic ambiguity resolution. *Psychological Review* 101(4), 676.
Mahowald, K., A. James., R. Futrell and E. Gibson. 2016. A meta-analysis of syntactic priming in language production. *Journal of Memory and Language* 91, 5-27.
McRae, K., M. J. Spivey-Knowlton and M. K. Tanenhaus. 1998. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language* 38(3), 283-312.
Michaelov, J. A., C. Arnett., T. A Chang and B. K. Bergen. 2023. Structural priming demonstrates abstract grammatical representations in multilingual language models. arXiv preprint arXiv:2311.09194.
Pickering, M. J. and H. P. Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39(4), 633-651.
Pickering, M. J. and V. S. Ferreira. 2008. Structural priming: A critical review. *Psychological Bulletin* 134(3), 427.
Prasad, G., M. Van Schijndel and T. Linzen. 2019. Using priming to uncover the organization of syntactic representations in neural language models. arXiv preprint arXiv:1909.10579.
Reitter, D., F. Keller and J. D. Moore. 2011. A computational cognitive model of syntactic priming. *Cognitive Science* 35(4), 587-637.
Scheepers, C. 2003. Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition* 89(3), 179-205.

© 2025 KASELL All rights reserved

Sinclair, A., J. Jumelet, W. Zuidema and R. Fernández. 2021. Syntactic persistence in language models: Priming as a window into abstract language representations. arXiv preprint arXiv:2109.14989.

Spivey-Knowlton, M. and J. C. Sedivy. 1995. Resolving attachment ambiguities with multiple constraints. *Cognition* 55(3), 227-267.

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

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

Tooley, K. M. 2023. Structural priming during comprehension: A pattern from many pieces. *Psychonomic Bulletin & Review* 30(3), 882-896.

Tree, J. E. F. and P. J. Meijer. 1999. Building syntactic structure in speaking. *Journal of Psycholinguistic Research* 28, 71-90.

Van Schijndel, M. and T. Linzen. 2018. A neural model of adaptation in reading. arXiv preprint arXiv:1808.09930.

Xu, W. and R. Futrell. 2024. A hierarchical Bayesian model for syntactic priming. arXiv preprint arXiv:2405.15964.

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

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

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary

© 2025 KASELL All rights reserved