Frequency-dependent Regularization in Constituent Ordering Preferences

Zoey Liu (yiliu@ucdavis.edu)

Emily Morgan (eimorgan@ucdavis.edu)
Department of Linguistics, University of California, Davis
One Shields Avenue, Davis, CA 95616 USA

Abstract

We examine how idiosyncrasies of specific verbs in syntactic constructions affect constituent ordering preferences. Previous work on binomial expressions in English has demonstrated that the polarization of ordering preferences for a given binomial type depends on its overall frequency. The higher the frequency of a binomial type, the stronger and more extreme preference/regularization language users will have for one alternative over the other (e.g. "facts and techniques" > "techniques and facts"; "bread and butter" >>> "butter and bread"). Here using the dative constructions in English as the test case, we show that the same frequency-dependent regularization exists in syntactic structures above the word level. The more frequent a dative construction type is, governed by the head verb, the stronger preference there is for one alternation over the other. Further, we present evidence that the regularization patterns can be accounted for via iterated learning modeling of language change, suggesting that frequency-dependent regularization emerges via the interactions between language production, language learning and cultural transmission.

Keywords: idiosyncratic preference; regularization bias; dative construction; iterated learning

Introduction

When a verb can have multiple syntactic subcategorization frames, argument structure realizations are governed by both abstract constraints (e.g. definite noun phrases tend to appear before indefinite ones) and item-specific knowledge (e.g. a particular verb has a bias towards different subcategorization frames) (Goldberg, 2003, 2009). Previous research on word order preferences has mainly examined effects of various abstract constraints. For instance, it has been shown extensively in both corpus studies and psycholinguistic experiments that syntactic structures have the tendency to shorten overall or average dependency length (Futrell, Mahowald, & Gibson, 2015), to prefer animate objects before inanimate objects (Christianson & Ferreira, 2005), or to put given discourse elements before new ones (Prat-Sala & Branigan, 2000).

On the other hand, other studies have noted the significant role of item-specific knowledge in predicting syntactic choices, reasoning that language users have idiosyncratic preferences in word orders (Morgan & Levy, 2016a). For example, the ordering preference for binomial expressions in English (X and Y) is largely affected by lexical, semantic and phonological properties of the words in the binomials (e.g. shorter word appears first; the final syllable of the second word should not be stressed). These abstract factors might predict that there is not a sharp discrepancy in the preference strength or extremity for *safe and sound* compared to that for *sound and safe*. Nevertheless, language users prevalently prefer *safe and sound* due to their more frequent usage of this particular order. This makes the ordering preference for this binomial type relatively more consistent and conventionalized.

Within the context of word order preference, the consistent preference for one structural variant among all syntactic alternatives (e.g. *safe and sound* is regularly more preferred to *sound and safe*) is known as *regularization*, a well-known phenomenon in statistical learning. In the linguistic domain in general, *regularization* refers to the tendency to make language structure more systematic and fixed, which minimizes the extent of variation in language usage. [1] Specifically in this paper, regularization refers to the phenomenon that when there is variation in the input, language users would preferentially reproduce the most frequent alternative that they have encountered. [2] For example, if a speaker hears the binomial type including *safe and sound* and *sound and safe*, the former will be more preferred in production due to its overall higher frequency in the input. Previous experiments in language learning and production have shown that both adults and children tend to regularize their output given the input (Hudson Kam & Newport, 2005), which can potentially explain why truly unpredictable or free linguistic variation is rare.

Nevertheless, the pressure to regularize contradicts the dominant view from research situated in rational language processing (Levy, 2008). This line of work posits with mounting evidence that language users are sensitive to the probabilistic distribution of different linguistic structures. Therefore they should perform *probability matching* rather than regularization. For a given structural type, they would reproduce all alternatives such that the ratios for these alternatives match their original probability in the input. For instance, if a speaker has encountered *safe and sound* 80 times and *sound and safe* 20 times (a 4:1 ratio), in their production the ratio for these two variants will approximate 4:1 as well. In this case, instead of minimizing variation, language

[1] The notion of *regularization* here differs from morphological regularity which refers to certain linguistic items abiding by compositional rules when going through morphological processes.

[2] Similar predictions are also made by the concept of *entrenchment* in the literature of cognitive grammar (Langacker, 1987).

users are likely to exhibit behaviors that maintain the same or similar distributions to the given input.

Morgan and Levy (2016b) demonstrated regularization in ordering preferences for binomial expressions in English. They showed that the higher the overall frequency of the binomial type is, the more regularized and extreme preference language users have for one alternative over the other (*radio and television* > *television and radio*; *salt and pepper* >>> *pepper and salt*). In other words, the extent of regularization is frequency-dependent.

They further demonstrated that it is possible to account for this frequency-dependent regularization bias with computational modeling of language change. Specifically, they adopted Iterated Learning Models (Reali & Griffiths, 2009) which allow us to simulate how language changes over generations. Though standard Iterated Learning Models are able to capture a general tendency for regularization, they cannot capture the relationship between the preference extremity and the frequency of an expression. To overcome this, Morgan and Levy (2016b) incorporated a frequency-independent regularization bias function during the data generation stage in their models. This leads to emergence of frequency-dependent regularization in the stationary distribution and the models are able to predict preference extremity for binomials as observed in corpus data.

As fruitful as previous findings are, most studies which have demonstrated a regularization bias have focused on learning and production of individual words or phrases (Hudson Kam & Newport, 2005), while explorations of regularization in syntactic constructions at a higher level are lacking. Thus in general, whether language users tend to perform probability matching or regularization when reproducing structural variants and under what context remain far from clear.

This study makes a contribution towards this gap. Following Morgan and Levy (2016b), we investigate the role of verb idiosyncrasy in constituent ordering preferences for abstract syntactic constructions above the word level. Leveraging large-scale corpus data, we address two questions. First, does the same frequency-dependent regularization found at the word level for binomials in English also operate on more complex syntactic levels? Secondly, how does this frequency-dependent regularization bias emerge?

In comparison to binomial expressions, the regularization pattern in more abstract syntactic structures might be different. The length of binomials is relatively short (in Morgan and Levy (2016b) all expressions have a length of 3 words) and it mostly involves orderings of two words (e.g. whether to put *safe* before *sound* or vice versa). By contrast, larger constituents in syntactic constructions above the word level tend to be much longer. Thereby deciding the relative order of larger constituents potentially involves more processing effort in comparison to binomials, which possibly leads to a stronger regularization bias (Ferdinand, Kirby, & Smith, 2019).

On the other hand, given that more abstract structures possibly contain more syntactic complexity than binomials, when ordering larger constituents, it might be the case that there are more lexical and structural constraints at different linguistic levels that should be taken into consideration. Accordingly, even if verbs have idiosyncratic preferences, they may not be exerting a significant effect. In this case we might see a weaker extent of regularization compared to that in binomials instead.

Our testbed

To address our questions, we use the dative construction (Bresnan, 2007; Bresnan & Ford, 2010; Yi, Koenig, & Roland, 2019) in English as the testbed. We define a dative verb as one that can appear in either the double object structure (V-NP-NP), as in (1), or the prepositional object structure (V-NP-PP), as in (2) (it does not need to appear in both structures). In this way different dative types are distinguished based on their head verbs.

(1) I **sent** [$_{NP}$ the reviewers] [$_{NP}$ the paper].

(2) I **gave** [$_{NP}$ the comments] [$_{PP}$ to the authors].

It is entirely true that not every verb identified by our criteria would traditionally be considered a dative verb. It is also possible that if a verb appears in the double object structure, its prepositional object alternative might not be considered *grammatical* under certain circumstances, and vice versa. Nevertheless, the motivations for our decisions are threefold. First, there is no definitive and concrete criteria to judge whether a verb is a dative verb or not. The seminal work of Levin (1993) offers a list of dative verbs in English, where verbs that can appear in only one structural alternation were also mentioned.

Secondly, as we are taking a data-driven approach from a usage-based perspective (Bybee, 1985; Cameron-Faulkner, Lieven, & Tomasello, 2003; Dabrowska, 2008; Ellis, 2002), we let corpus observations decide the syntactic properties of a verb. For instance, we regard a verb as ditransitive if it appears in a ditransitive/double object construction, even it might not have been previously considered as a representative case in the literature.

Thirdly, here we focus on the idea that if a verb appears in a double object structure, it has the potential to appear in a prepositional object order, even it might be considered ungrammatical under certain contexts. Similarly, if a verb appears in a prepositional object structure, we consider it as also having the potential to occur in a double object order, regardless of whether it is always deemed grammatical. The main reason is that the notion of grammaticality varies among different language users under different conditions. One crucial point noted in previous experiments (Bresnan, Cueni, Nikitina, & Baayen, 2007) is that not every alternation headed by a verb has to have a grammatical alternative, because whether there is a grammatical syntactic alternation is constrained by abstract factors such as dependency length and pronominality. To illustrate this, let us consider the following examples. The verb *give* is perhaps one of the most typical dative verb, yet many consider (4) as ungrammatical because Budapest is not an animate object, and accordingly it can not serve as the recipient of the action.

(3) She gave the draft to him.

(4) She gave the draft to Budapest.

Nevertheless, the grammaticality of (4) might be regarded differently in a particular context. The sentence can be perfectly fine if it is used as a metonym (e.g. referring to the publisher office located in Budapest) (Hovav & Levin, 2008).

With the dative construction, we explore whether there is a relationship between regularization/preference extremity and the overall frequency of each dative type. We further investigate the origin of the regularization bias with computational modeling of language change, exploring in particular whether Iterated Learning Models can predict frequency-dependent regularization.

Existence of frequency-dependent regularization

Overall construction frequency

For estimates of the overall frequency for each dative construction type, we used raw data in English from the CoNLL 2017 Shared Task (Ginter, Hajič, Luotolahti, Straka, & Zeman, 2017) on multilingual parsing, which has a total of around 9 billion tokens. This corpus consists of web page data from both Common Crawl and Wikipedia and is automatically parsed with UDPipe (Straka & Straková, 2017). Within this corpus, each token is represented in an individual line with its morphosyntactic information encoded, including parts-of-speech tags and syntactic dependency relations. For double object structures, we extracted sentences in which the head verb takes one direct object and one indirect object. For prepositional object structures, we extracted sentences in which the head verb takes one direct object and one PP oblique with the functional head *to* that follows the direct object.

In all these cases, the head verbs are restricted to only lexical verbs (not including auxiliaries). We treated each verb as a dative type, then calculated the overall frequency of each type. Dative types with a frequency lower than 1000 in total were removed to ensure that we have enough data to reliably estimate the frequency of different alternations. After preprocessing, our dataset contains 733 unique dative types (733 unique head verbs) with a total of 13 million dative instances. Among these, 4 million appear in the double object structure while 9 million appear in the prepositional object structure.

Preference extremity

Our goal is to estimate the role of verb idiosyncrasy in preference extremity for each dative construction type. Here the preference extremity should be approximated as the probability for the more preferred structure between the two alternations within each construction type. Since argument structure realizations for specific instances of the dative constructions are constrained by abstract factors (e.g. phrasal length), we have to exclude the effects of these factors in order to more accurately quantify the influence of verb idiosyncrasy.

To do this, we first fit a mixed-effect logistic regression model to predict the prepositional object order in our dataset following Bresnan et al. (2007). We included verb as a random effect and included fixed effects for three automatically measurable factors: definiteness, pronominality and phrasal length. Under an ideal circumstance, the model would contain other factors that have been tested in previous dative alternation studies such as animacy and the semantic class of the head verb. However, those factors require manual coding, which is not realistic given our settings.

Since we are interested in how specific verbs affect ordering preference extremity, we estimated verb bias (Stallings, MacDonald, & O'Seaghdha, 1998; Wasow & Arnold, 2003) for each verb as the probability of a sentence being realized as the prepositional object structure based on just the random effect of the verb (eliminating the contributions from the fixed effects). Specifically, let V be the random effect intercept of a particular verb derived from the regression model, the probability of this verb being realized in the V-NP-PP structure is then calculated as follows.

$$\frac{1}{1+exp(-1*V)} \qquad (1)$$

Given each dative type, a probability value larger than 0.50 indicates that the prepositional object structure is preferred over the double object order, and the preference extremity is the same as the probability value. When the probability is lower than 0.50, this indicates a stronger preference for the double object order, and the preference extremity is computed as the absolute difference between this probability value and 1.

After collecting both the overall frequency and preference extremity of each dative type, we fit a linear regression with the former being the predictor and the latter being the outcome variable in order to evaluate the significance of overall frequency.

Results

As presented in Figure 1, frequency-dependent regularization does exist in the dative constructions ($\beta = 0.04, p < 0.01$). The higher the overall frequency of the dative type is (denoted by the head verb), the stronger the preference extremity is for one alternation over the other. Among the most extreme cases are the typical dative verbs such as *give*, *bring*, *send*, which all favor double object order as predicted by the logistic regression model.

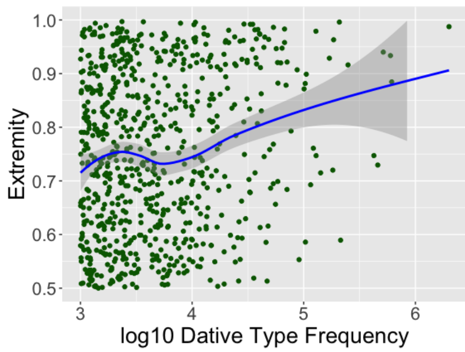

Figure 1: Plot of preference extremity against log10 overall construction frequency for 733 dative construction types. Overall frequency is estimated from corpus data for English from Ginter et al. (2017).

Accounting for frequency-dependent regularization

Now that we have seen that frequency-dependent regularization exists in the dative construction, we turn to the second question raised at the beginning: how does frequency-dependent regularization occur in the first place? To address this, we borrow Iterated Learning Models (hereafter ILM), which are computational models that simulate language change. We first introduce standard 2-alternative ILMs. Since standard ILMs do not encode the relationship between regularization and overall frequency, we also introduce the augmented ILMs from Morgan and Levy (2016b). Their models were able to predict frequency-dependent regularization via applying a regularization function during the data generation stage. Our models follow Morgan and Levy (2016b). We describe the simulation procedures of our models for the corpus data and show that they successfully predict frequency-dependent regularization in the dative construction.

Standard ILMs

Iterated learning has gained wide popularity over recent years as an approach to study how language evolves through cultural transmission. The crucial insight of this methodology is that language structures are transmitted culturally via language users learning those structures from others' usage patterns of the same structures. Meanwhile during each stage of transmission and learning, language learners can impose their own biases on the usage of the structures as well, which in turn shapes language structures.

ILMs computationally simulate this learning process, where the output of the previous learner is fed as the input to the next learner and this process proceeds in an iterative fashion. If the tendency to regularize emerges from reproduction of structural alternatives based on the input, which is also a process that continues iteratively, ILMs serve as an ideal tool to account for regularization.

For each dative construction type, the basic idea for the learning process is as follows. Imagine a learner hears all N instances of a dative type as the input, with there being x in the prepositional object order, and $N-x$ in the double object order. The learner infers a hypothesis θ_1 of the probability that the dative type appears as a prepositional object structure based on the input, then produces new data. The next learner repeats the same procedures.

The prior probability of the dative type being in the prepositional object order (the probability of θ_1) is expressed as drawn from the beta distribution (B) with two parameters: μ and ν. The former defines the mean of the distribution while the latter determines the width or the concentration of the distribution.

$$P(\theta_1) = \frac{\theta_1^{\mu\nu-1}(1-\theta_1)^{(1-\mu)\nu-1}}{B(\mu\nu,(1-\mu)\nu)} \tag{2}$$

As for interpretation, μ represents the ordering preference for a given dative construction type. The learners in all generations are set to have the same μ in the learning process of each dative type. As we assume learners have no innate knowledge of which structure will be more preferred and what their relative probabilities should be, we assign μ with a value of 0.5. This means that the learner believes the two alternative structures have equal probability of occurrence.

On the other hand, ν reflects the confidence in the prior probability, with a higher value representing the learner is fairly confident about their prior knowledge, and a lower value representing vice versa. Different from μ, ν is a free parameter. When μ is 0.5, as shown in Figure 2, a higher value of ν indicates that the prior probability distribution centers around 0.5. This means that most of the time the learner believes the prepositional object structure and the double object structure will appear for roughly equal proportions of times, which corresponds to more structural variation. A lower value of ν denotes that the prior probability distribution is more scattered. This means that most of the time the learner believes within each dative type, one alternation is more preferred than the other, which leads to more regularization.

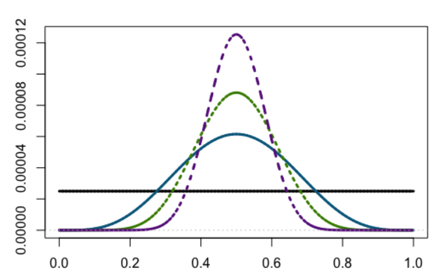

Figure 2: Plot of prior probability distribution when μ is 0.50 with different ν values: ν = 2 (black, which is the uniform distribution); ν = 10 (blue); ν = 20 (green); ν = 40 (purple).

Given a learner's hypothesis θ_1, new data is generated following the binomial distribution.

$$P(x|\theta_1) = \binom{N}{x}\theta_1^x(1-\theta_1)^{N-x_1} \tag{3}$$

The next learner applies Bayes rule to calculate a posterior distribution over all hypotheses of the dative type being realized as the prepositional object order. Since the Beta distribution is the conjugate prior of the binomial distribution, the posterior also follows a Beta distribution. The learner then samples a hypothesis θ_1 from this posterior distribution and continues to generate data.

ILMs from Morgan and Levy (2016b)

Nevertheless, as previously mentioned, standard ILMs do not predict frequency-dependent regularization. To solve this, Morgan and Levy (2016b) augmented standard ILMs at the stage of data generation. Within each generation of an augmented model, the learner applies a regularization function with a free bias parameter R to update θ_1 and form a new hypothesis θ_1'. The regularization function itself is frequency-independent, in the sense that the value for R holds in each stage for all dative types. A higher value of R corresponds to more pressure to regularize. The model then generates data based on θ_1'.

$$\theta_1' = \frac{\theta_1^R}{\theta_1^R + (1-\theta_1)^R} \tag{4}$$

Simulating corpus data

To predict the frequency-dependent regularization observed in our corpus data of the dative construction, we followed Morgan and Levy (2016b) and introduced a frequency-independent regularization function with a free parameter R at each data generation stage. In the simulation process, we set N of each dative construction type to be an approximate for the number of times a college student who is a native speaker of English has been exposed to that particular dative type. The estimate for the lifetime linguistic exposure of a college-age native English speaker is around 300 million words in total (Levy, Fedorenko, Breen, & Gibson, 2012). For each dative type, we ran 50 chains of learners for 1500 generations. This is not to suggest that realistically the process of language learning and production has continued for 1500 generations, but rather in order for the model's learning process to reach the stationary distribution. Within each chain, θ_1 is initialized as 0.50. For hypothesis updating, we experimented with a series of different values for the two free parameters ν and R ($\nu = \{2,3,4,5,6,7,8\}$; $R = \{1,1.1,1.3,1.5,1.8,2,2.1,2.3,2.5,2.8,3,3.1,3.3,3.5,3.8,4\}$), resulting in a total of 112 models. We collected θ_1' from the final generation of each chain. A θ_1' value higher than 0.50 indicates a preference for the prepositional object structure over the double object structure, and the predicted preference extremity by the model is the same as θ_1'. If θ_1' is smaller than 0.50, the double object structure is the more preferred order between the two alternatives and the predicted preference strength is measured as 1 - θ_1'.

Results

If ILMs are able to account for regularization, as the overall frequency of the construction type increases, the value for predicted preference extremity should increase as well. Results from Figure 3 corroborate our findings in Figure 1, showing that we can predict frequency-dependent regularization in the dative construction with combinations of appropriate values for ν and R, though to different extents (Table 1).

We see the most comparable patterns to that in Figure 1 when ν equals 2. This lends support to our motivation of initializing the prior probability as 0.50, since a Beta distribution with μ of 0.50 and ν of 2 is the uniform distribution, which means that our prior is truly uninformative, i.e. learners have no innate knowledge of which structure is more preferred. When R value is held constant, the pressure to regularize is weaker as ν increases.

Our observations here differ from Morgan and Levy (2016b) in one aspect. They demonstrated that regularization in binomial expressions in English already emerges in their models when R is as low as 1.1, yet with much larger values for ν ($\nu = \{10,15,20\}$). Recall that a lower value for R as well as a higher value for ν both correspond to a weaker regularization bias. This means that the extent of regularization is stronger in the dative construction than that in binomials. We return to this point in the Discussion section.

Table 1: Linear regression (predicting preference extremity as a function of overall frequency) results for subgraphs in Figure 3.

	$\nu = 2$	$\nu = 4$
$R = 1$	$(\beta = 0.00, p = 0.11)$	$(\beta = 0.00, p = 0.60)$
$R = 2$	$(\beta = 0.01, p < 0.01)$	$(\beta = 0.01, p < 0.01)$
$R = 3$	$(\beta = 0.02, p < 0.01)$	$(\beta = 0.02, p < 0.01)$
$R = 4$	$(\beta = 0.03, p < 0.01)$	$(\beta = 0.02, p < 0.01)$

Discussion

Using the dative construction in English as the test case, we have demonstrated frequency-dependent regularization in constituent ordering preferences in abstract syntactic constructions above the word level. The more frequent the construction type is, governed by the head verb, the more polarized preference language users have for one syntactic variant over the other. In addition, the second question we have addressed is regarding the origin of frequency-dependent regularization. Recall that while standard ILM is not able to account for frequency-dependent regularization, the results have shown that when combined with a frequency-independent regularization bias, the augmented model is able to predict the observed regularization patterns in the dative constructions. This indicates that just language processing

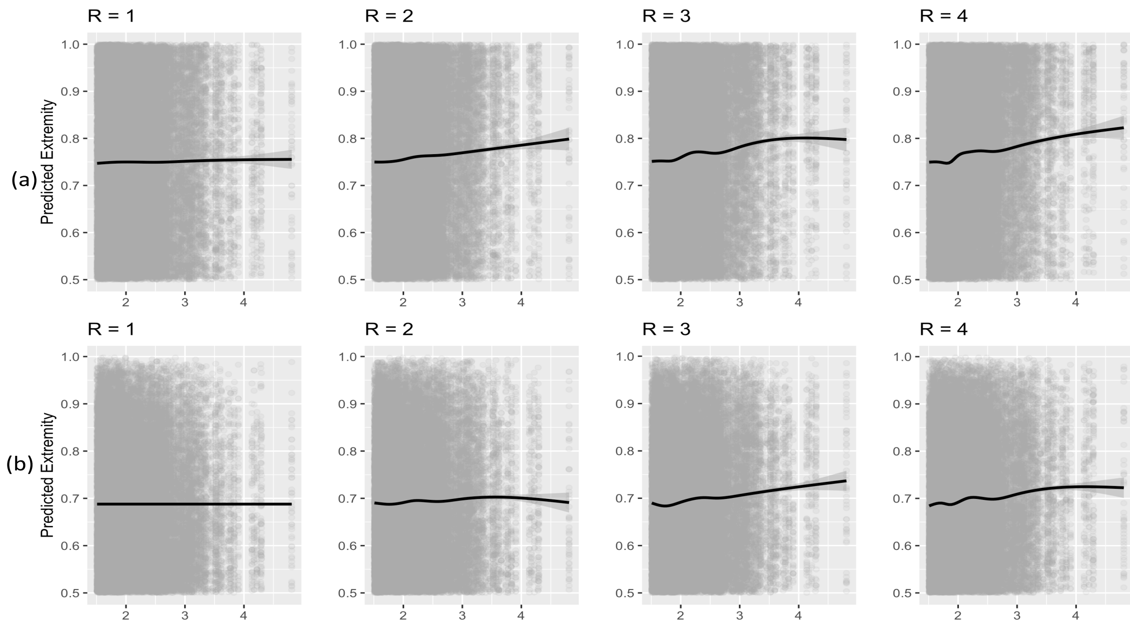

Figure 3: Selected plots of predicted extremity values from ILM with different values of *R* when: ν = 2 ((a)); ν = 4 ((b)).

alone is not enough to yield frequency-dependent regularization, but rather this pattern arises from the continuous interactions between language production as well as the process of cultural transmission and language learning.

Contrary to our study, previous experiments on verb idiosyncrasy have mainly focused on comprehension tasks rather than the production (Garnsey, Pearlmutter, Myers, & Lotocky, 1997; MacDonald, 1994). They have demonstrated that comprehenders perform probability matching and that the probabilistic information of verb subcategorization frames is able to predict processing behaviors. For example, the verb *suggest* has a stronger preference for taking a sentential complement (e.g. *We* ***suggest*** *that it is time to investigate verb bias.*), rather than having a direct object (e.g. *The reviewers* ***suggest*** *more examples.*). Accordingly, a sentence where *suggest* is followed by a sentential complement is proportionally easier to process compared to one where *suggest* takes a direct object.

Overall, all these findings indicate that a cohesive and complete account for language structures and processing patterns should incorporate item-specific knowledge along with abstract factors. Indeed, Morgan and Levy (2015) have presented that the model which has the best performance in predicting the distribution of preference polarization for binomial expressions in English is the one that takes into account both abstract constraints as well as the overall frequency of the binomial types. Results from comprehension tasks in Morgan and Levy (2016a) also showed that online processing patterns of highly frequent binomials are directly shaped by their frequency.

One remaining question is why there is a stronger regularization bias in the dative construction compared to binomials. Previous work has shown that the extent of regularization depends on cognitive load (Ferdinand et al., 2019). Learners tend to regularize more when the cognitive load needed by the specific learning tasks is high. Comparing binomials and the dative constructions, the two have different levels of syntactic complexity. The ordering of a binomial expression only involves the two content words within the binomial. For instance, with *safe and sound*, a language user mainly needs to figure out whether to put *safe* or *sound* first. This is relatively much easier than the argument realization of a dative construction, where a language user has to decide whether to use a V-NP-PP order, or a V-NP-NP order, both of which have many more words and more nested hierarchical structures. Since the dative construction is structurally more complex, its ordering might involve more cognitive load than ordering the two content words in a binomial, which results in language users having more regularized ordering preferences.

Further experiments on idiosyncrasy in other types of syntactic alternations such as adjective ordering or adverb placement, especially in a crosslinguistic context, would provide valuable insights into the existence and extent of regularization. Methodologically, the ILM that we have adopted here assumes that one learner only takes the input of one other learner, whereas in reality language users learn from multiple sources at the same time. Smith et al. (2017) successfully approximated the learning process via letting the learner takes input from more than one speaker within each generation. Future work should explore how different model types compare in their explanatory power of regularization.

References

Bresnan, J. (2007). Is syntactic knowledge probabilistic?

Experiments with the English dative alternation. *Roots: Linguistics in search of its evidential base*, *96*, 77–96.

Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In *Cognitive foundations of interpretation* (pp. 69–94). KNAW.

Bresnan, J., & Ford, M. (2010). Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language*, *86*(1), 168–213.

Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form* (Vol. 9). John Benjamins Publishing.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child directed speech. *Cognitive Science*, *27*(6), 843–873.

Christianson, K., & Ferreira, F. (2005). Conceptual accessibility and sentence production in a free word order language (Odawa). *Cognition*, *98*(2), 105–135.

Dabrowska, E. (2008). Questions with long-distance dependencies: A usage-based perspective. *Cognitive Linguistics*, *19*(3), 391–425.

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition*, *24*(2), 143–188.

Ferdinand, V., Kirby, S., & Smith, K. (2019). The cognitive roots of regularization in language. *Cognition*, *184*, 53–68.

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336–10341.

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, *37*(1), 58–93.

Ginter, F., Hajič, J., Luotolahti, J., Straka, M., & Zeman, D. (2017). *CoNLL 2017 Shared Task - Automatically Annotated Raw Texts and Word Embeddings.* (LINDAT/CLARIN digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University)

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences*, *7*(5), 219–224.

Goldberg, A. E. (2009). The nature of generalization in language. *Cognitive Linguistics*, *20*(1), 93–127.

Hovav, M. R., & Levin, B. (2008). The English dative alternation: The case for verb sensitivity. *Journal of linguistics*, *44*(1), 129–167.

Hudson Kam, C. L., & Newport, E. L. (2005). Regularizing Unpredictable Variation: The Roles of Adult and Child Learners in Language Formation and Change. *Language Learning and Development*, *1*(2), 151-195.

Langacker, R. W. (1987). *Foundations of cognitive grammar: Theoretical prerequisites* (Vol. 1). Stanford university press.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126-1177.

Levy, R., Fedorenko, E., Breen, M., & Gibson, E. (2012). The processing of extraposed structures in English. *Cognition*, *122*(1), 12–36.

MacDonald, M. C. (1994). Probabilistic constraints and syntactic ambiguity resolution. *Language and Cognitive Processes*, *9*(2), 157-201.

Morgan, E., & Levy, R. (2015). Modeling idiosyncratic preferences: How generative knowledge and expression frequency jointly determine language structure. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society* (p. 1649–1654).

Morgan, E., & Levy, R. (2016a). Abstract knowledge versus direct experience in processing of binomial expressions. *Cognition*, *157*, 384 - 402.

Morgan, E., & Levy, R. (2016b). Frequency-Dependent Regularization in Iterated Learning. In *The Evolution of Language: Proceedings of the 11th International Conference (evolang11).*

Prat-Sala, M., & Branigan, H. P. (2000). Discourse constraints on syntactic processing in language production: A cross-linguistic study in English and Spanish. *Journal of Memory and Language*, *42*(2), 168–182.

Reali, F., & Griffiths, T. L. (2009). The evolution of frequency distributions: Relating regularization to inductive biases through iterated learning. *Cognition*, *111*(3), 317–328.

Smith, K., Perfors, A., Fehér, O., Samara, A., Swoboda, K., & Wonnacott, E. (2017). Language learning, language use and the evolution of linguistic variation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1711), 20160051.

Stallings, L. M., MacDonald, M. C., & O'Seaghdha, P. G. (1998). Phrasal ordering constraints in sentence production: Phrase length and verb disposition in heavy-NP shift. *Journal of Memory and Language*, *39*(3), 392–417.

Straka, M., & Straková, J. (2017, August). Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe. In *Proceedings of the CoNLL 2017 shared task: Multilingual parsing from raw text to universal dependencies* (pp. 88–99). Vancouver, Canada: Association for Computational Linguistics.

Wasow, T., & Arnold, J. (2003). Post-verbal constituent ordering in English. *Topics in English Linguistics*, *43*, 119–154.

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