

# Optimal Auxiliary Selection in English: A View from Gradient Harmonic Grammar

Kazuma KIMURA

Graduate student, University of Tsukuba, Japan  
s2320038@u.tsukuba.ac.jp

**Abstract:** This paper investigates a more fine-grained explanation of the phenomena BE-perfect in English from the viewpoint of Gradient Harmonic Grammar, which is a stochastic version of constraint-based grammar (e.g., Optimality Theory). It is known that in perfect constructions the auxiliary is altered between BE and HAVE, which is called Auxiliary Selection. However, the alternation of auxiliaries is cross-linguistically not categorial but gradient. BE-perfect in English is a typical example, where the adjunction of a specific PP slightly degrades the acceptability of BE. In order to address the gradeability in BE-perfect, we propose the combination the constraint-based evaluation with a probabilistic model employing a *likelihood function*. In this model, the grammaticality and acceptability of linguistic expressions are assessed probabilistically, ranging from 0 to 1. The probability is computed based on the degree of violations of constraints, referred to as *Harmony*. We will show in this paper that the proposed model predicts gradient acceptability of BE-perfect sentences in a successful manner. In addition to this, we will argue the interplay between the stochastic model (GHG) and sentence processing theory. This study will be a first step towards a principled computational model of human sentence processing.

**Keywords:** Gradient Harmonic Grammar, auxiliary selection, BE-perfect, English

## 1. Introduction

There has been argued that like other naturalistic phenomena linguistic phenomena generally shows probabilistic nature rather than categorial one across broader domain such as syntax, semantics, and phonology (Pierrehumbert (2003), Kitahara (2009), Hayes (2022)). To capture such a probabilistic graduality, in the linguistic literature, several models of stochastic Optimality Theory (e.g., Stochastic OT, Harmonic Grammar (HG), or Maximum Entropy (ME)) have been proposed, each of which are all constraint-based models. Within the framework of OT, grammaticality and acceptability hinge on the violation of linguistic constraints, akin to grammatical rules. Of all the existing models just mentioned, “Gradient Harmonic Grammar” (GHG) (Smolensky and Legendre (2006), Pater (2016)) stands out, eschewing strictly dominated constraints<sup>[1]</sup>. GHG evaluates a degree called *Harmony* in terms of weight assigned to constraint violations and linguistic symbols. The definition of Harmony is as follows (adopted from Pater (2009)).

$$(1) \quad \begin{array}{ll} \textit{Harmony} & S_k = \text{weight of a linguistic symbol} \\ H(x) = \sum S_k w_k & w_k = \text{weight of a constraint } C_l \end{array}$$

In GHG, unlike traditional OT approaches, each constraint is weighted by numeral degree. Suppose that  $S_k$  represents the weight of a linguistic symbol (any features relevant to linguistic representation) and  $w_l$  denotes the violation weight of  $C_l$ . They are assigned the weight 2.0 and 1.5 respectively. The harmony of an input  $x$  (including  $S_k$  and  $w_k$ ) is evaluated by calculating the product of the weight of  $S_k$  and  $C_l$ 's violation weight (resulting in  $2.0 \times 1.5 = 3.0$ ). This model is notably fine-grained than traditional OT model in that it allows a gradient (numerical) output. Consequently, it enables a more nuanced (non-categorial) representation of grammaticality across various domains such as syntax, semantics, and phonology.

As Lee and Müller (2018) points out, however, there are few GHG studies which focus on syntactic phenomena or morphosyntactic ones, though with respect to phonological phenomena a

handful of proposals have been put forth (see, Boersma and Hayes (2001), Goldwater and Johnson (2003), Jager (2003), Boersma and Pater (2008)). To substantiate the validity of the model, an examination of the morphosyntactic domain (i.e., the interface between syntax and morphology) becomes imperative. The GHG model employed here then targets a morphosyntactic phenomenon known as *Auxiliary Selection*, primarily observed in Romance languages.

In general, several romance languages such as French, Dutch, and Italian show the gradation of auxiliary selection, depending on the semantic properties of verbal domain. Numerous studies have been endeavored to submit a unified and principled explanation of this gradation from various perspectives (syntax, Kayne (1993), Bjorkman (2011), Fujita (2016), Pietraszko (2023); semantics, Sorace (2000), Washio (2004), Legendre (2007); its interface, Baker (2019), Amato (2022)). One of the most prominent studies is Baker (2019): this study claims that the selection of auxiliary relies on functional heads with specific semantic properties (e.g., change, telicity, or control) based on a purely semantic analysis proposed by Sorace (2000). While such a syntactic approach captures a preference of auxiliary selection to some extent, it encounters challenges in explaining the nuanced nature of auxiliary insertion, particularly in cases where both auxiliaries BE and HAVE are deemed acceptable.

In this paper, our attention is directed towards a specific case of BE-perfect in English. Although BE-perfect is scarcely tolerated in the modern English, some exceptional cases (e.g., *is gone*, *is finished*) are observed. More important, when a specific prepositional phrase (PP) is adjoined, the acceptability is highly degraded. Such a phenomenon is characterized as showing gradual acceptability. Thus, the goal of this paper is to provide a GHG analysis of the nuanced acceptability, which is observed in BE-perfect in English. Given that GHG are required to be applicable to morphosyntactic phenomena and BE-perfect sentences are challenging for traditional approaches, we thus expect that this study contributes to both (i) validating the GHG model and (ii) explaining a particular phenomenon altogether.

The organization of this paper is as follows. Section 2 reviews the basic facts about BE-perfect in the modern English and then introduces Baker’s (2019) proposal, which is the base of our proposal. In section 3, we present the architecture of GHG. In section 4, which is the main part of this paper, we give our proposal and the result of GHG evaluation. Section 5 discusses further implications of the proposal, focusing on the interaction with sentence processing theory. Section 6 concludes this paper.

## 2. Backgrounds: facts and theory

### 2.1 Basic facts: exceptional BE-perfect in English

Although there has been argued that auxiliary selection disappeared in the modern English, several expectational cases, where the auxiliary BE is selected, are still observed. (2a) is a standard perfect auxiliary, where HAVE is selected. (2b), however, is also an acceptable sentence with auxiliary BE selected. Interestingly, both sentences are interpreted and acceptable as purely perfec(ive) construction.

- (2) a. John had gone.  
 b. John was gone.
- (3) a. John has gone to the store.  
 b. ?John is gone to the store.  
 c. John has gone on the interstate.  
 d. ?\*John had gone on the interstate.  
 e. The fight is finished. (Lispon (1999: 56))

Note, however, that the exceptional cases in (3b, d) are slightly degraded when a directional PP (hereafter, to-PP) or path PP (hereafter, on-PP) is adjoined to the BE-perfect. Additionally, the case of on-PP has a slight lower acceptability than that of to-PP. This suggests that the selection of auxiliary heavily depends on the aspectual property of verbal domain, since these PPs are assumed to change the aspectual interpretation of verb.

### 2.2 Functional projection and auxiliary selection

Given that auxiliary selection is subject to aspectual property of verbal domain, several syntactic analyses have been proposed from the viewpoint of the interaction between syntactic structure and aspectual interpretation. Despite the abundance of numerous prior studies adopting a syntactico-semantic approach to auxiliary selection (Kayne (1993), Bjorkman (2011) and others), we will restrict our focus to Baker (2019), which bears the utmost relevance to our discussion, for the sake of argumentative coherence.

Baker (2019) argues that a multi-layered functional structure as in (4) is in a certain correlational relationship with the selection of auxiliary. Each functional head includes an eventive feature which corresponds to its semantic property: control head=[+control], Initiation head = [+initiation], state head = [+state], change head = [+change], and telic head = [+telic]. [+control] and [+initiation] is attributed to control of an action by the agent; [+state] the static property of an action; and [+change] and [+telic] change of state/location).

(4) [ControlP **Control** [InitiationP **Initiation** [StateP **State** [ChangeP **Change** [TelicP **Telic** [VP ... ]]]]]]

Baker (2019) claims that in several languages such as French, Dutch, German, and Italian the selection of auxiliary is dependent on these semantic features as Table 1 shows<sup>[2]</sup>. The bottom of the table shows that [+telic] and [+change] (change of state) prefer to inserting BE. In contrast, [+control] and [+initiation] (e.g., Controlled process) prefer to inserting HAVE as shown in the upper part of the table.

Table1

*Auxiliary Selection and Functional Structure (Baker (2019:585))*

	French	Dutch	German	Italian
<b>Controlled process (non-motional)</b>	HAVE	HAVE	HAVE	(BE)/HAVE
<b>Controlled process (motional)</b>	(BE?)/HAVE	BE/HAVE	(BE)/HAVE	(BE)/HAVE
Uncontrolled process	HAVE	HAVE	HAVE	(BE)/HAVE
Existence of state	HAVE	(BE)/HAVE	(BE)/HAVE	BE/(HAVE)
Continuation of a pre-existing state	HAVE	BE/HAVE	BE/HAVE	BE/(HAVE)
<b>Change of state</b>	BE/HAVE	BE/(HAVE)	BE	BE/(HAVE)
Change of location	BE	BE	BE	BE

The same analysis may be apparently applicable to the examples of English BE-perfect in (2b) since the BE-perfects receive change of state or telic interpretation (the termination of John's going somewhere). Following the Baker's proposal, the syntactic representations of (3b) and (3d) are the following:

(5) a. [ Control [Initiation [State [Change to-PP<sub>i</sub> [Telic t<sub>i</sub> ]]]]]]  
 b. [ Control [Initiation [State on-PP [Change [Telic ]]]]]]

### 2.3 Issues

So far, we have reviewed the exceptional cases where auxiliary BE is selected (BE-perfect) and the cartographic syntactic analysis proposed by Baker (2019). Interestingly, assuming the functional structure in (4) is able to capture the distribution of each auxiliary, with BE being preferred to [+Telic], [+change], and HAVE [+Control], [+Initiation]. There are, however, remaining questions: (a) why does cross-linguistic variation (e.g., gradual optionality or obligatoriness) exist; (b) how the acceptability contrast between (3b) and (3d) should be accounted for?

To give appropriate answers to these questions, we explore a more fine-grained probabilistic model, namely GHG, which is a novel mechanism of generating various nuanced linguistic expressions. Unfortunately, we do not deal with the first question since answering the question requires a huge cross-linguistic corpus analysis and acceptability survey, both of which exceed the scope of this study. Therefore, our attention turns to answering the second question by using GHG. The GHG model is

based on a constraint-based architecture as assumed in OT, and therefore capable of capturing the gradient acceptability such as (2b, d). In the preceding sections, the core basics of GHG model (section3) and the results of GHG evaluation (section 4) will be given.

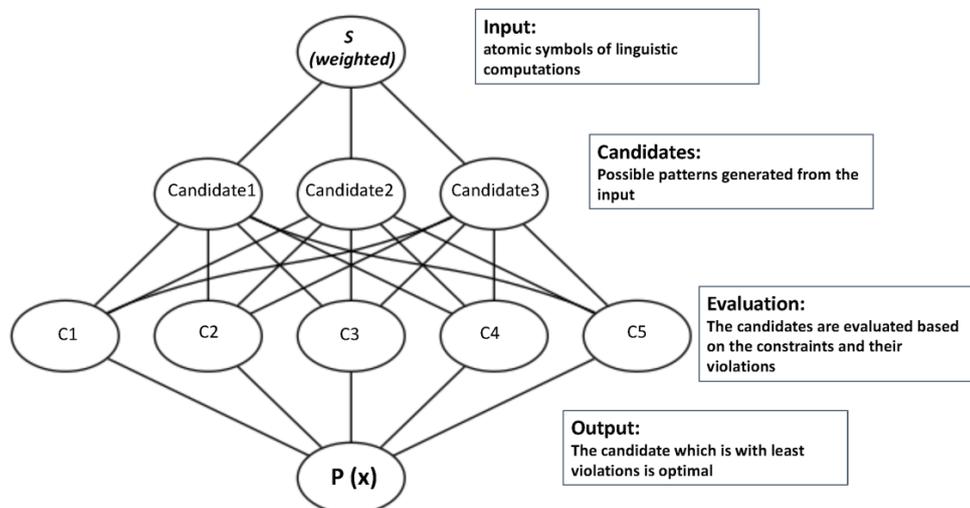
### 3. GHG architecture

This section introduces a basic architecture of GHG. Within the stochastic constrained-based framework of grammar such as GHG, grammatical outputs are generally evaluated by means of *Optimality*, which of the concept is defined as shown below:

- (6) *Optimality*  
 An output qualifies as optimal if it is the candidate with maximal harmony in its candidate set.  
 A candidate has maximal harmony if it has the value closet to zero (i.e. the lowest penalty).  
 (Lee and Müller (2018))

The optimality of linguistic inputs is determined in terms of the violation of grammatical constraint as traditional OT approach does. In addition to this, stochastic OT assumes that the results of evaluation are transformed into the form of probability (the value from 0 to 1) as the likelihood function (log-transformation). The GHG architecture is given below.

Figure 1  
 The GHG architecture



The input S (linguistic symbols) are transformed into three possible candidates (Candidate 1 ~ Candidate 3) and are evaluated by weighted constraints (C1~C5). This pattern is the same as the traditional OT architecture. More important, within GHG, it is not only the constraints but also linguistic inputs (symbols) that are assigned weights (Smolensky and Goldrick (2016)). Linguistic symbols here include any element relevant to linguistic representation: abstract formal features in syntax and morphology, semantic features, phonological distinctive features and so on. This stipulation suggests that both constraints and linguistic symbols are non-categorical and gradient. Thus, Figure 1 above shows that the input nodes, and the constraint nodes are weighted respectively.

The GHG architecture requires the candidates evaluated by weighted constraints to be transformed into the form of possibility ( $P(x)$ ). The probability is calculated by likelihood functions, which are defined as Equations (1) and (2).

$$\begin{aligned} \text{The likelihood function (adopted from Culbertson et al (2013))} \\ \log(P(x|G)) + z = H_G(x) = -\sum_k w_k S_k & \quad (1) \\ P(x|G) = e^{-\sum_k w_k S_k(x)} & \quad (2) \end{aligned}$$

The first equation denotes that the Harmony of a linguistic expression given a grammar  $G$  is represented as the sum of the product of a constraint and its violation weight. The second equation means that the sum of violation constraints is transformed by an exponential function and the output ranges from 0 to 1. This transformation is similar with the sigmoid function, which is used in the field of Natural Language Processing or Deep Learning as in Figure 2. In the context of GHG, y-axis corresponds to the final probability of a linguistic expression, and x-axis the baseline Harmony  $H_G(x)$ . As suggested in the right side of Figure 2, the higher score the Baseline Harmony is, the higher the probability is.

Let us consider a simple example of GHG evaluation. Suppose that two inputs  $x_i$  and  $x_j$ . Both candidates are evaluated in terms of the violations of several constraints. Subsequently, the final probability of a linguistic expression  $x_i$  is calculated based on the evaluation score of each candidate. The equation is given below.

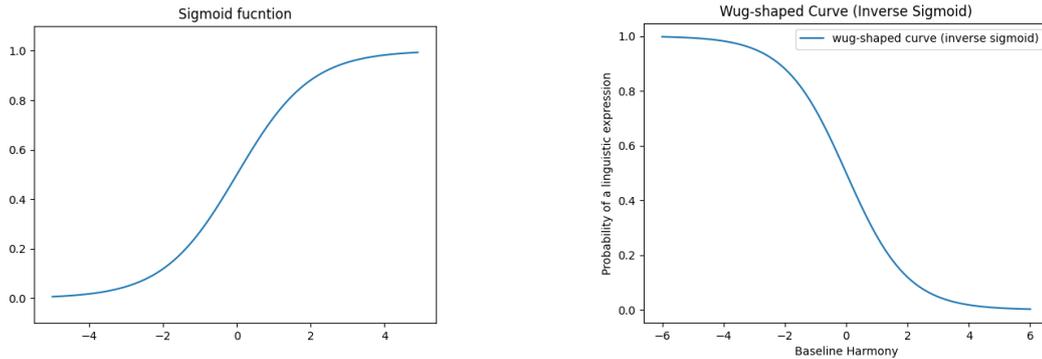
$$P(x_i | G) = e^{-\sum_k w_k S_k(x_i)} / Z, \text{ where } Z = e^{-\sum_k w_k S_k(x_i)} + e^{-\sum_k w_k S_k(x_j)} \quad (3)$$

$Z$  is the sum of the values of each candidate calculated by the exponential function. If the Harmonies of  $x_i$  and  $x_j$  are 3 and 2.5 respectively, then the probability of  $x_i$  is computed in the form of  $P(x_i | G) = e^{-3} / Z$  ( $Z = e^{-3} + e^{-2.5}$ ). Since  $e^{-3}$  and  $e^{-2.5}$  are 0.0498 and 0.0821. The probabilities of  $x_i$  and  $x_j$  amount to about 37% and 63%. Notice that the closer to 0 Harmony value is the higher possibility the candidate amounts to.

It is worth noting that such a probabilistic evaluation captures a non-categorical phenomena like nuanced acceptability judgement in (3b) and (3d) from a geometric viewpoint. The probability transformed by the exponential function exhibits a gradual curve as the sigmoid function explored in the context of machine learning, in contrast to step functions and similar, which have abrupt change (the value is either 0 or 1).

Figure 2

*Sigmoid function and wug-shaped curve (adopted from Hayes (2022: 482), with slight modifications)*



Given that the insertion of such a gradient function layer into neural network computation experiences a significant leap in performance, it is not unreasonable to conclude that our linguistic judgement can be derived from a gradient function. Thus, we adopt GHG in order to provide an explanation of the acceptability contrast between (3d) and (3d). In the following section, we present our proposal using GHG and give a stochastic analysis.

## 4. Proposal

### 4.1. Externalization conditions as constraints

Based on the Baker’s analysis reviewed in section2, we define the relationship between functional heads and inserted auxiliaries as the following set of constraints:

- (7) a. \*Con  $\Leftrightarrow$  BE (Control head does not select BE)
- b. \*Init  $\Leftrightarrow$  BE (Initiation head does not select BE)

- c. \*St  $\Leftrightarrow$  BE (State head does not select BE)
- d. \*Ch  $\Leftrightarrow$  HAVE (Change head does not select HAVE)
- e. \*Te  $\Leftrightarrow$  HAVE (Telic head does not select HAVE)

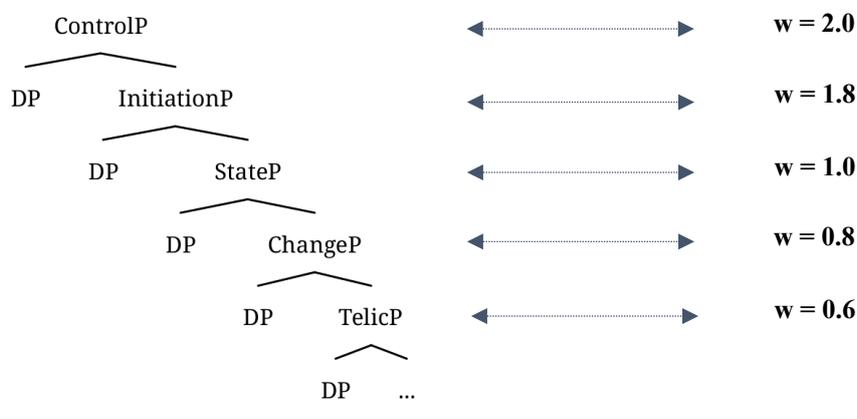
Recall that verbs with [+control] and [+initiation] tend to select HAVE and verbs with [+change] and [+telic] BE. To put it differently, these tendencies dictate the prohibition from selecting BE in an environment where [+control] and [+initiation] are activated, or from selecting HAVE in an environment where [+change] and [+telic] are activated. Furthermore, since GHG requires that each constraint be assigned the weights respectively, we assume the weights of the five constraints:

- (8)
- a. \*Con  $\Leftrightarrow$  BE [2.0]
  - b. \*Init  $\Leftrightarrow$  BE [1.8]
  - c. \*St  $\Leftrightarrow$  BE [1.2]
  - d. \*Ch  $\Leftrightarrow$  HAVE [0.8]
  - e. \*Te  $\Leftrightarrow$  HAVE [0.6]

Notice that the constraints which restrict the selection of BE (=8a-c) are assigned higher values than the constraints restricting the selection of HAVE (= (8d, e)). The hierarchy of this weighing is not a stipulation without any motivation; rather it is determined based on syntactic structure. As shown in Figure 3, the weights of each constraint parallels to the functional hierarchical structure:

Figure 3

*The relationship between hierarchical structure and weights of constraints*



This implies that the functional heads in a structurally higher position bear a strong burden to linguistic computation (here GHG evaluation) while the functional heads in a lower position are considered to have lower computational load. We will discuss the issue with weighting of constraints and its relationship to syntactic structure (syntactic processing) in section 5.

In addition, the GHG model we adopt assumes that the auxiliaries BE and HAVE (linguistic symbols) are assigned the weights respectively. This is motivated by the fact that there is a lexical difference between BE and HAVE: BE is inherently a copular verb and HAVE a possession verb (Kayne (1993), Myler (2016)). Thus, we give the weights to the auxiliaries as follows.

- (9)
- a. BE [1.0]
  - b. HAVE [1.5]

The difference in the assigned weight is assumed to reflect the lexical property of each auxiliary: HAVE is lexically stronger than BE for some reasons, which will be out of the scope of this paper (for more detailed discussion of this issue, see Bravo et al (2015)). Since GHG takes into consideration the sum of constraint weights  $S_k$  and their violation weight  $w_k$ , these weights are regarded as constraint weights and are calculated in part of Harmony  $H_G(x)$ .

## 4.2 GHG evaluation

We are now in a position to calculate the Harmonies of syntactic representations of BE-perfect in English. First, we should discuss the syntactic structures of (2a-d). In the case of directional PP, namely (2a, b), the PP is generated in the Spec-TelicP and Sepc-ChangeP as shown in (8a). This is because directional PPs are interpreted as the “termination” of an event. In contrast, (8b), where the path PP is adjoined, requires the PP to be generated in the Spec-StateP due to the semantic property of the PP. Path PPs are generally interpreted as activating the “process” of an event.

- (10) a. [ Control<sub>[2.0]</sub> [Initiation<sub>[1.8]</sub> [State<sub>[1.2]</sub> [Change<sub>[0.8]</sub> to-PP<sub>i</sub> [Telic<sub>[0.6]</sub> t<sub>i</sub> ]]]]]  
 b. [ Control<sub>[2.0]</sub> [Initiation<sub>[1.8]</sub> [State<sub>[1.2]</sub> on-PP [Change<sub>[0.4]</sub> [Telic<sub>[0.6]</sub> ]]]]]

In addition, it is assumed that the auxiliaries BE and HAVE are inserted after syntactic computation has completed along the line of Distributed Morphology.

The probabilities of (8a, b) are calculated in terms of the likelihood function (Harmony  $H_G(x)$  and  $P(x|G)$ ) discussed above. First, in (8a), we find that TelicP and ChangeP are activated since the to-PP is merged with their Spec positions. Since this representation is interpreted as a perfect construction, ControlP, InitiationP, and StateP are also activated. If HAVE is inserted into this representation, then, \*Tel and \*Ch are violated. On the other hand, the insertion of BE violates \*Cont, \*Init, and \*St. The candidate 1 (inserting HAVE) violates \*Ch and \*Te and the candidate 2 (inserting BE) violates \*Con and \*Init. Thus, the Harmonies are calculated as follows ( $x_1$  = inserting HAVE into (8a) and  $x_2$  = inserting BE into (8a)):

- (11) a.  $H(x_1) = - \{ (1.5 \times 0.6) + (1.5 \times 0.8) \} = -2.1$   
 b.  $H(x_2) = - \{ (1.0 \times 2.0) + (1.0 \times 1.8) \} = -3.8$

Based on the harmony values -2.1 and -3.8, the final probabilities  $P(x)$  is calculated as  $P(x_1 | G) = e^{-2.1} / Z$ , and  $e^{-3.8} / Z$  where  $e^{-2.1} = 0.1221$ ,  $e^{-3.8} = 0.0211$ , and  $Z = 0.1432$ . Thus, the probabilities of each candidate are 85.3% and 14.7% respectively.

Table 2  
The evaluation of (10a)

Input:	Violations	H(x)	P(x)
1. (8a)⇔HAVE	*Ch, *Tel	-2.1	0.8526...
2. (8a)⇔BE	*Con, *Init	<b>-3.8</b>	<b>0.1473...</b>

Once again, we see that the harmony value closer to zero (here -2.1) yields higher probability.

The Harmonies and the probabilities in (8b) are calculated in the same manner. However, the case where BE is inserted violates \*St since the path PP is base-generated in Spec-StateP. Thus, the Harmony of the candidate 2 is -5.0 ( $-\{ (1.0 \times 2.0) + (1.0 \times 1.8) + (1.0 \times 1.2) \}$ ), so the sum of the exponential Harmonies in (8b) amounts to 0.1288. The results are summarized below.

Table 3  
The evaluation of (10b)

Input:	Violations	H(x)	P(x)
1. (8b)⇔HAVE	*Ch, *Tel	-2.1	0.9479...
2. (8b) ⇔ BE	*Con, *Init, *St	<b>-5.0</b>	<b>0.0520...</b>

Notice that in both cases the probability of HAVE-perfect is higher than that of BE-perfect (85% vs 14% in (10a), 94% vs 5% in (10b)). Comparison of Table 2 and Table 3 leads us to an expected conclusion that the adjunction of on-PP sharply degrades its acceptability: in the case of (10a), the probability of BE-perfect amounts to 14%, which is relatively higher than that of (10b). It is compatible with the data in (3), where the acceptability of (3b) sentences slightly degrades, but that of (3d) sharply degrades.

We provided the GHG analysis of BE-perfect sentences, in which the adjunction of a certain PP degraded the acceptability of BE. The nuanced acceptability of BE-perfect sentences is derived from the interaction between the weights assigned to the auxiliaries and the violation weights of constraints in (8). Furthermore, the fact that on-PP BE-perfect is less acceptable than to-PP BE-perfect is due to the violation of \*St ([+state] does not select auxiliary BE). The evaluations by the likelihood function reveal that this analysis predicts the distribution of acceptability in a successful way<sup>[3]</sup>.

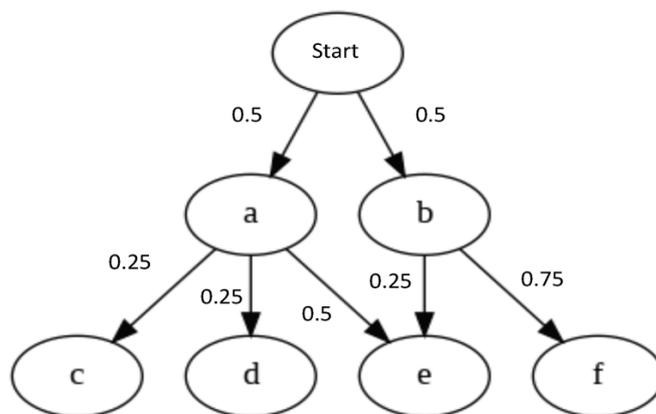
## 5. Further issue: A psycholinguistic view

In this section, our focus turns to a remaining issue with the contrast between BE and HAVE mentioned in section 3. We assumed that there is a difference in the value assigned to BE (1.0) and HAVE (1.5). The question immediately arises: why such a weighting is adopted? To tackle with the issue, we expand our view to the processing of BE and HAVE sentences. Although we do not give a principled answer, an implication is discussed from the view of the interplay between GHG and sentence processing. As mentioned above, GHG assumes that linguistic symbols are weighted as linguistic constraints. We hypothesize that the weighting difference between BE and HAVE comes from the real time sentence processing load, which is assessed by *Competition hypothesis* and *Entropy reduction* (Linzen and Jaeger (2014)).

A simple schematic example of these hypotheses is given in Figure 4. The grammar assumed here is expected to generate two possible output strings: *a-e* or *b-e*. The probabilities of the first string *a* and *b* are 0.5. The node *a* has several options of its incoming element *c*, *d*, or *e* while the node *b* has two options *e* and *f*. Note that each node is assigned the probability of being selected.

Figure 4

*Example of a mini grammar assessed by Competition hypothesis and Entropy reduction*  
(Linzen and Jaeger (2014), with slight modifications)



Of the two assessments of sentence processing mentioned above, the former one *Competition hypothesis* predicts the next element based on the entropy of subcategorization. Compared the *a* node and the *b* node, the entropy of the *a*'s subcategorization is higher than that of the *b*'s subcategorization: The *a* node activates three possible next nodes while the *b* node activates only two nodes. Consequently, this hypothesis predicts that longer reading time of *a* node is required. On the other hand, the other hypothesis *Entropy reduction*, under which the reading time of a node is assessed in terms of the difference among the probabilities of incoming elements (the reduction of Entropy). In the schematic

example above, the reading time of *b* node is longer since the degree of *Entropy reduction* is higher in the *e* node and *f* node ( $0.75 - 0.25 = 0.5$ ) than the *e* node and *c* or *d* node ( $0.5 - 0.25 = 0.25$ ). Although Linzen and Jaeger (2014) tentatively concludes that *Competition hypothesis* reflects the sentence processing of human, it is still controversial which hypothesis is a better model of human language processing.

Let us consider the real time processing of BE and HAVE sentences as shown in below.

- (12) a. The man had gone to the store.  
b. The man was gone to the store.

When we read up to the auxiliary verbs, we generally predict the next syntactic structure (subcategorization frame) since we are uncertain what element (e.g., VP, NP, clause structure or PP) at the point of reading the auxiliaries. It is a well-known fact that both BE and HAVE selects various of complements. BE is involved with copular NP and AP, PP, past/present participle while HAVE requires a possessed NP, small clause (SC), past participle. Suppose the case where BE requires more possible subcategories than HAVE. Under *Competition hypothesis*, the entropy value of BE-subcategorization is higher than that of HAVE. Thus, the reading time on BE is longer due to computational burden imposed on accessing the subcategories. In contrast, *Entropy reduction* hypothesis predicts incoming elements on the point where we process BE and HAVE based on the difference in probabilities of their complements. As for the reduction of their entropy is unclear. Regardless of which hypothesis to adopt, however, it is assumed that BE and HAVE lexically encode different information. This issue, of course, will not be solved until we conduct a self-paced reading experiment which targets on the BE and HAVE sentences and computational implementation (e.g., Probability Context Free Grammar, PCFG).

## 6. Concluding remarks

We proposed an analysis of BE-perfect based on GHG: the subtle acceptability differences introduced by certain PPs are explained by assuming that the values of weights imposed on the inserted auxiliaries and those imposed on the constraints itself interact with each other and are probabilistically calculated in terms of the likelihood function. On the other hand, this analysis did not provide a principled explanation for the assignment of different values to BE and HAVE, specifically that HAVE is assigned a higher weight. Therefore, in section 5, we presented a consideration of GHG weighting from the perspective of sentence processing metrics. Given the result of this study and the implication from it, two future tasks are as follows: (a) verification of cross-linguistic data (e.g., French, Dutch, German, and Italian) and (b) conducting self-paced reading (SPR) experiment in which the target is BE and HAVE sentences in English.

## Notes

- [1] As Kakeru Yazawa (personal communication) points out, there are several stochastic models of such a grammatical system other than GHG (e.g., Maximum Entropy model, or Noisy Harmonic Grammar), each of which adopt a similar computational architecture. In the morphosyntactic literature, however, GHG is a common method to evaluate apparently categorial phenomena (e.g., word order, or the realization of morpheme) in a non-categorial way. This leads us to the (tentative) selection of GHG in this paper.
- [2] Under Baker's proposal, it might be that both auxiliaries are post-syntactically inserted into somewhere of the functional head although Baker does not state the position of auxiliary's realization. If so, however, this approach will not require the distinction between main verbs and auxiliaries. For simplicity this paper does not refer to the syntactic position of auxiliaries.
- [3] A potential issue with this analysis is the weighting issue: why is such a weight assigning allowed? To deal with this issue, we need to conduct experimental research (e.g., a corpus study or acceptability judgement task). The method of such research needs to be considered in our future work.

## Acknowledgements

My special gratitude goes to Shuto Yamamura, Yuichi Ono, Ryohei Naya, Hiroyuki Iwasaki, and Kakeru Yazawa for their insightful comments on this study. I am also indebted to Akitaka Yamada and Sogo Kawaguchi for their valuable feedback on the preliminary study poster presented at the 167<sup>th</sup> meeting of Linguistic Society of Japan held at Doshisha University. My special thanks to Takuto Kimura and Yuki Okada for their collaboration of the preliminary study presented at the 44<sup>th</sup> conference of Tsukuba English Linguistics Society (Kimura, Okada, and Kimura (in prep)). All errors and shortcomings that may remain are entirely my own.

## References

- Amato, I. (2022). Auxiliary Selection is Agree: person-driven and argument-structure-based splits, *Isogloss*, 8(2), 1-20.
- Baker, J. (2019). Split Intransitivity in English, *English Language and Linguistics*, 23(3), 557-589.
- Bjorkman, B. (2011). BE-ing default: the morphosyntax of auxiliaries, Doctoral Dissertation, Massachusetts Institute of Technology.
- Boersma, P. & B. Hayes. (2001). Empirical tests of the gradual learning algorithm, *Linguistic Inquiry*, 32(1). 45-86.
- Brennan, J., S. Dingare., & C. Manning (2001). Soft constraints mirror hard constraints: Voice and person in English and Lummi. In Miriam Butt & Tracy Holloway King (Eds), *Proceedings of the LFG 01 conference*. Stanford, CA: CSLI Publications.
- Culbertson, J., P. Smolensky, & C. Wilson. (2013) Cognitive Biases, Linguistic Universals, and constraint-based Grammar Learning, *Topics in Cognitive Science*, 5, 394-424.
- Goldwater, S. & M. Johnson. (2003) Learning OT constraint rankings using a Maximum Entropy model. In J. Spencer., A. Ericsson., & O. Dahl (Eds), *Proceedings of the Stockholm Workshop on Variation Within Optimality Theory*, 113-22. Stockholm: Stockholm University.
- Hale, K. & Keyser, S. J. (1993) On argument structure and the lexical expression of syntactic relations. In Kenneth Hale & Samuel Jay Keyser (Eds), *The view from Building 20*, 53-110. Cambridge, MA: MIT Press.
- Hayes, B. (2022) Deriving the wug-shaped curve: A criterion for assessing formal theories of linguistic variation, *The Annual Review of Linguistics*, 8, 473-94.
- Lee, H., & Müller, G. (2018). Asymmetries in long-distance dependencies: a view from GradientHarmonic Grammar. Paper presented at the Workshop on Long-Distance Dependencies, Humboldt-Universität zu Berlin, October 4-5.
- Legendre, Geraldine. (2007). On the typology of auxiliary selection in Romance. *Lingua*, 117(9), 1522-40.
- Linzen, T. & T. F. Jaeger. (2014). Investigating the role of entropy in sentence processing, *Proceedings of the 2014 ACL Workshop on Cognitive Modeling and Computational Linguistics*, 10-18.
- Lipson, M. (1999). The loss of Auxiliary Selection in English. *University of Pennsylvania Working Papers in Linguistics*, 6(2), 54-67.
- Sorace, A. (2000). Gradients in auxiliary selection with intransitive verbs. *Language*, 76(4), 859-90.
- Pater, J. (2009). Weighted constraints in generative linguistics. *Cognitive Science*, 33, 999-1035.
- Pater, J. (2016). Universal Grammar with Weighted Constraints. In J. McCarthy & J. Pater (Eds), *Harmonic Grammar and Harmonic Serialism*, 1-46, Equinox: Sheffield.
- Pietraszko, A. (2023). Cyclic selection: Auxiliaries are merged, not inserted. *Linguistic Inquiry*, 54(2), 350-377.
- Smolensky, P. & G. Legendre. (2006). *The harmonic mind*. Cambridge, MA: MIT Press.
- Smolensky, P. & M. Goldrick. (2016). Gradient symbolic representations in grammar: The case of French Liaison. *Rutgers Optimality Archive* (ROA) 1286.