EVALUATING TWO GRAMMAR TYPES FOR JAPANESE: A PRELIMINARY STUDY[*]

Tomohiro Fujii and Hibiki Yamashita
Yokohama National University

1. Introduction

The present paper concerns grammar evaluation. It is always an issue of interest for theoretical syntacticians what syntactic structures should be assigned to sentences in a language and what set of rules is needed as the grammar of the language in order to generate the structures. Sometimes, evidence is available in favor of one set of rules or another. It is also often the case, however, that data can be analyzed in different ways and therefore more than one grammar is compatible with the data. The issue of an *evaluation metric* arises here (Chomsky 1957, 1965; Chomsky and Halle 1968). In this paper, we introduce an exercise question, where two grammar types for Japanese compete, and discuss what evaluation metric can be thought of in order to resolve the particular theory selection problem. (It will be clear why we call the question an "exercise.")

In the past decades, it has been debated how head-final surface sentence forms are generated in Japanese and other similar languages like Korean. Just to name a few: Aoyagi 2006; Farmer 1980; Fujii 2016; Fukui and Sakai 2003; Fukui and Takano 2000; Funakoshi 2014, 2017, 2020; Han, Lidz, and Musolino 2007, 2015; Hayashi and Fujii 2015, 2016; Kayne 1994; Kishimoto 2008; Kitagawa 1986; Koizumi 2000; Koopman 2005; Kuroda 1965, 2003; Miyagawa 1980, 1999; Murasugi 2000; Otani and Whitman 1991; Saito 2012; Saito and Fukui 1998; Saito and Hoshi 1998, 2000; Sato and Hayashi 2018; Sells 1995; Shibata 2015; Shimada 2007; Takita 2009, 2013; Tokizaki and Kuwana 2013; Tonoike 1995, Whitman 1998, among others.

Fujii (2016) examines the structures of subordinated clauses headed by complementizers. The complement clause found in (3) is an example.

(1) Hiroshi-wa DeNA-ga katta-to omotta.
Hiroshi-TOP DeNA-NOM won-C thought

'Hiroshi thought that the DeNA BayStars had won.'

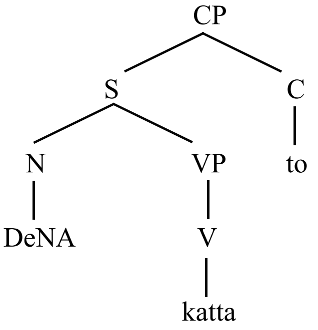
Fujii applies standard constituency tests, which reveal that the sting *DeNA-ga katta-to* is assigned a structure like the one in (2), as opposed to the one in (3), suggesting that Japanese native speakers internalize what we might call a "Complement-Head Grammar" in (2), not what

---

[*] This work is partially supported by JSPS KAKENHI Grant Number JP 19K00569.

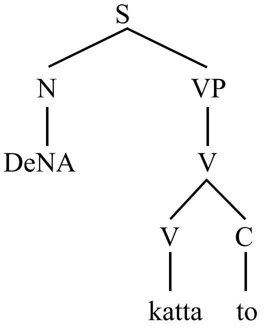
we might call a "Head-Head Grammar" in (3).[1] (We abstract away from the existence of Tense and multi-word nominal strings for the ease of presentation.)

(2)



A Complement-Head Grammar
CP → S C
S → N VP
VP → V | N V
N → DeNA | SoftBank
V → katta
C → to

(3)



A Head-Head Grammar
S → N VP
VP → V | N V
V → V C
N → DeNA | SoftBank
V → katta
C → to

Fujii discusses some pieces of evidence in favor the CH Grammar. Take one based on a kind of proform replacement. In (4), the indeterminate proform *nan(i)* 'what' replaces the substring that is exhaustively dominated by S. The HH Grammar predicts that this kind of proform replacement would be impossible because the string *DeNA-ga katta* is not a constituent under the grammar.

(4) A: Hiroshi-wa nan-to iimasita-ka?
Hiroshi-TOP what-C said-Q
'What did Hiroshi say?'

B: DeNA-ga katta-to iimasita.
DeNA-NOM won-C said
'He said that DeNA had won.'

It can thus be concluded that the CH Grammar is descriptively superior to the HH Grammar.

This said, let us introduce the question that concerns us. Suppose we set aside empirical

---

[1] See Sells (1995) for a concrete proposal that falls within the realm of HH Grammars. See also Kitagawa (1986), Shimada (2007), and Saito (2012). These authors propose that head movement takes place out of head adjunction structure in one way or another.

arguments for one alternative like the one mentioned above. Then, is there any way to decide if either the CH Grammar or HH Grammar more highly valued than the other? Recall that we have called our question an exercise question. There are a couple of reasons for doing so. First, even though arguments like one based on (4) exist, we discuss the two hypotheses as if these arguments did not exist. Second, we also know that even the CH Grammar is hard to maintain. Many grammatical components of Japanese are ignored here: argument drop, scrambling, wh-in-situ questions, case marking, and so on and so forth. Simple context-free grammars like the one in (2) cannot deal with them at all. We however set aside this fact and limit our scope to context-free grammars. Third, the two kinds of grammars both assume the non-terminal symbols S and VP. It should be noted that the literature has shown that these non-terminal symbols are not well-motivated. These category labels are problematic under X-bar Theory (Chomsky 1970, Stowell 1981) and/or Bare-Phrase-Structure Theory (Chomsky 1995). We do not find lexical items of the category S. Nor does the Inclusiveness Condition allow us to rely on bar-level notations, although they may be needed in order to cover certain facts. Thus, we know both grammars are problematic before we compare them.

Despite these reasons to believe the setting is not as realistic as it may be, we still think that these simple grammars are worth studying. Yang (2017) observes that “[f]rom the 1970s on, [. . .] the Evaluation Metric all but vanished except in the relative small field of formal learnability. There are multiple reasons for this, some reasonable and some not, but the net effect is that a mechanistic approach to language, strongly embodied in early generative grammar, received less attention”. While we agree with Yang’s observation that evaluation metrics have rarely been discussed for a while, we think that simplicity, parsimony, and non-redundancy are still considered as significant measures to evaluate theories with in minimalist theorizing too (Chomsky 1995, Freidin and Lasnik 2011). If so, it might be useful to carefully examine a simple and somewhat hypothetical case like ours to see how it is possible to measure the values of individual grammars.

In what follows, we propose to use the Bayesian model of grammar acquisition advanced by Perfors, Tenenbaum, and Regier (2006, 2011) (hereafter PTR) in order to examines the two grammar types for Japanese mentioned above. PTR’s central concern is in how children could choose “hierarchical grammars” especially over “flat grammars” without direct evidence. The former grammars are context-free grammars generating structured tree diagrams like (5a) whereas the latter are those generating ones like (5b). PTR show that their evaluation metric for rewriting grammars successfully choose hierarchical grammars over other grammars including flat grammars. Although the model obviously is intended for grammar acquisition, it can also be viewed as an evaluation metric for theory selection (Chomsky 1965, Yang 2017, Pearl 2017).

(5) a. [$_{\text{S}}$ [$_{\text{NP}}$ [$_{\text{Det}}$ the] [$_{\text{N}}$ boy]] [$_{\text{VP}}$ [V admires] [$_{\text{NP}}$ [$_{\text{N}}$ her]]]]

b. [$_{\text{S}}$ [$_{\text{Det}}$ the] [$_{\text{N}}$ boy] [$_{\text{V}}$ admires] [$_{\text{N}}$ her]]

As alluded to earlier, PTR's model utilizes Bayesian inference. This fact is not primarily important for our current purposes (see Yang 2017, Pearl 2020, and references cited there).[2] The major advantage of us starting with their model lies in the fact that the relationship between a grammar, data and the grammar's value are quite transparent. It is relatively easy to identify what revision of a grammar makes the grammar score better than the other.

The paper is structured as follows. Section 2 introduces and illustrates PTR's model with a concrete example. Section 3 attempts to use the model to compare CH and HH Grammars. Although we see that CH Grammars are regarded more highly valued than HH Grammars, we address several problems and argue that the results need to be interpreted with care. Proposing a way out of the problems with the first attempt, we make a second attempt to redo the comparison and critically examine the new results too. Section 4 concludes the paper.

2. The Model

In this section, we illustrate the model with two sets of context-free grammars that differ in a certain way in order to see how the model captures linguists' intuitions about simplicity and generalizability of grammars.

PTR's model, in accordance with Bayes' rule, argues that the posterior probability for a grammar G given a corpus D is proportional to the product of the prior probability for G and the likelihood for G on D, as in (6).

(6) $p(G|D) \propto p(D|G)p(G)$

The components of this rule are reviewed below.

2.1. Grammars and Their Prior Probabilities

In PTR's proposal, grammars are built in such a way that they have the minimum number of rules needed to generate the sentences in a given corpus. (7) is an example of a corpus. We call it "Corpus A." Corpora consist of strings of category symbols, not actual words, because PTR's study is primarily interested in acquisition of sentence patterns rather than lexical acquisition. The corpus includes *N V* and *Det N V*, as opposed to *John came* and *the girl ran*.

(7) *Corpus A*
N V
Det N V

This means that what we normally understand as *pre-terminal symbols* in context free grammars are treated as *terminals* here; see Kornai and Pullum (1990) for virtually the same approach.

---

[2] See Perfors, Tenenbaum, Griffiths, and Xu (2011), where a tutorial introduction to Bayesian models of cognitive development is provided.

Accordingly, let us introduce grammars that capture the corpus. There are two grammar types. We call one type "Always-NP Grammars" and the other "Sometimes-NP Grammars." The grammars differ in terms of how to conceive the notion "Noun Phrase". Under Always-NP Grammars, not only the string *N* but also the string *Det N* is an NP. Under Sometimes-NP Grammars, in contrast, the string *Det N* forms an NP while *N* does not. If we consider the smallest grammar of each type that generates the sentences, we obtain the sets of rules shown in Table 1.

Note also that PTR's system takes into consideration the number of non-terminals, that of terminals, and that of production rules in each grammar, when calculating prior probabilities. As shown in Table 1, Always-NP and Sometimes-NP Grammars are no different in terms of the numbers of the symbols and rules.

Table 1 The Always-NP Grammar and the Sometimes-NP Grammar built to cover the sentences in Corpus A

	The Always-NP Grammar		The Sometimes-NP Grammar	
Rules	S → NP V		S → N V | NP V	
	NP → Det N | N		NP → Det N	
Properties	# non-terminals	2	# non-terminals	2
	# terminals	3	# terminals	3
	# rules	3	# rules	3

With this setting in mind, let us see how the prior probabilities for the two grammars are calculated. The prior probability for a grammar *G* is calculated by the equation given in (8), where *V*, *n*, *P* and N_i indicate the number of vocabulary items, that of non-terminals, that of production rules, and that of right-hand-side items for production rule *i* respectively.[3]

(8) *Prior probability*

$$p(G) = p(P)p(n)\prod_{i=1}^{P} p(N_i)\prod_{j=1}^{N_i}\frac{1}{V}$$

Thus, for instance, *p(Always-NP Grammar)* reflects the following probabilities: (i) how likely a grammar ends up having 3 production rules, (ii) how likely a grammar ends up having 2 non-terminal symbols, (iii) how likely each production has the number of right-hand-side items that it has (e.g., *S →NP V* has two), and (iv) how likely each right-hand-side item for that production rule is chosen from among the five category symbols (e.g., *NP* in *S → NP V* is chosen with probability 1/5). Finally, PTR assume that *p(P)*, *p(n)*, and $p(N_i)$ are geometric distributions with parameter 0.5.[4] The larger the number of vocabulary symbols, the number of production rules, and the number of right-hand-side items of each production are, the lower *p(P)*, *p(n)*, and $p(N_i)$

---

[3] In addition, Perfors, Tanenbaum, and Regier (2011) take into account probabilities of production rules. We omit production probabilities in our calculations, following Perfors, Tanenbaum, and Regier (2006).

[4] Therefore, $p(n) = (1-0.5)^{n-1} * 0.5$.

are. Because we have to deal with very small numbers, the calculations are done in the log domain. (9) shows the prior probabilities for the two grammars that capture Corpus A.

(9) Log (p(Always-NP Grammar for Corpus A)) = -6.5051

Log (p(Sometimes-NP Grammar for Corpus A)) = -7.5051

The Always-NP Grammar is slightly preferred to the Sometimes-NP Grammar. We will see in Section 2.3 that a more remarkable difference emerges when the corpus is expanded.

2.2. Likelihood Probabilities for Grammars Given a Corpus

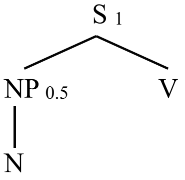
The likelihood probability for a grammar on a corpus is the probability of all the sentences in the corpus occurring at the same time under the grammar. In (10), the corpus *D* contains *k* sentences: $S_1, \ldots, S_k$.

(10) *Likelihood*

$$p(D|G) = \prod_{i=1}^{k}(S_i|G)$$

For instance, *N V* is parsed as in the tree diagram in (11) under the Always-NP Grammar. In this grammar, an S node is always rewritten this way because it has only one S rule. An NP node, in contrast, is rewritten as *Det N* 50 % of the time and as *N* 50 % of the time if the two N-rules are weighted equally. Thus, p([$_S$ [$_{NP}$ N] V]) = 1*0.5.

(11)



Production rule	Probability
S → NP V	1
NP → Det N	0.5
NP → N	0.5

Generally speaking, if there are many ways to rewrite a given non-terminal node under G, it contributes to decreasing the likelihood probability for G. Likewise, if a sentence is assigned a more complex tree (i.e., one containing more non-terminal nodes), it may also contribute to decreasing the likelihood.

Table 2. How the sentences of Corpus A are assigned probabilities by the Always-NP and Sometimes-NP Grammars.

	Sent. *S*	Tree *T* for *S*	# Ss in *T*	# NPs in *T*	Probability assigned to *S*
Always-NP G for Corpus A	N V	[$_S$ [$_{NP}$ N] V]	1	1	1*0.5 = 0.5
	Det N V	[$_S$ [$_{NP}$ Det N] V]	1	1	1*0.5 = 0.5
Sometimes-NP G for Corpus A	N V	[$_S$ N V]	1	0	0.5 = 0.5
	Det N V	[$_S$ [$_{NP}$ Det N] V]	1	1	0.5*1 = 0.5

The log prior, likelihood, and posterior probabilities for Always-NP and Sometimes-NP

Grammars for Corpus A are given in Table 3. The posterior for the Always-NP Grammar is slightly higher than that for the Sometimes-NP Grammar. This is due to the preference in prior for the former grammar over the latter because the two grammars' likelihoods are no different.

Table 3. Log prior, likelihood, and posterior probabilities for Always-NP and Sometimes-NP Grammars for Corpus A. The higher probabilities are highlighted.

Always-NP Grammar for Corpus A			Sometimes-NP Grammar for Corpus A		
Prior	Likelihood	Posterior	Prior	Likelihood	Posterior
-6.5051	-0.6020	-7.1072	-7.5051	-0.6020	-8.1072

2.3. Expanding the Corpus and Grammars' Generalizability

Two grammars that are compatible with a corpus may differ in such a way that one captures a linguistically significant generalization while the other misses it. In the case of grammar comparison discussed so far, Always-NP Grammars capture the generalization that the syntactic distributions of *N* and *Det N* are essentially the same. The string *the man* occurs where *John* occurs, and vice versa. We quickly see how the model captures this.

Let us extend Corpus A to include some new sentences. The larger corpus, Corpus B, has four new sentences, underscored in (12). We now have transitive sentences in our corpus. Accordingly, the Always-NP and Sometimes-NP Grammars for Corpus A are revised to cover the new data while keeping their major features the same: both *N* and *Det N* are dominated by *NP* in the revised Always-NP Grammar, whereas only *Det N* is dominated by *NP* in the revised Sometimes-NP Grammar.

(12) *Corpus B*

N V	N V N	Det N V N
Det N V	N V Det N	Det N V Det N

This extension of the corpus results in the following situation: while only one rule needs to be added in Always-NP Grammars, four rules need to be in Sometimes-NP Grammars.

Table 4. The rules and the numbers of the symbols/ rules of the Always-NP and Sometimes-NP Grammars for Corpus B.

	Always-NP Grammar for Corpus B		Sometimes-NP Grammar for Corpus B	
Rules	S → NP V | NP V NP NP → Det N | N		S → N V | NP V | N V N | N V NP NP V N | NP V NP NP → Det N	
# symbols and rules	# non-terminals	2	# non-terminals	2
	# terminals	3	# terminals	3
	# rules	4	# rules	7

The probabilities for the revised grammars are given in the second row in Table 5.

Table 5. Log prior, likelihood, and posterior probabilities for Always-NP and Sometimes-NP Grammars for Corpora A and B. For each corpus, the higher prior, likelihood, and posterior probabilities are highlighted.

	Always-NP Grammars			Sometimes-NP Grammars		
	Prior	Likelihood	Posterior	Prior	Likelihood	Posterior
For Corpus A	-6.5051	-0.6020	-7.1072	-7.5051	-0.6020	-8.1072
For Corpus B	-9.8061	-4.8164	-14.6226	-24.7092	-4.6689	-29.3781

Compared to what happens before the corpus gets extended, the posterior for the Always-NP Grammar is more clearly preferred to that for the Sometimes-NP Grammar. More specifically, what contributes to this difference most is that the Sometimes-NP Grammar's prior probability radically drops from -7.5051 to -24.7092. The prior drops in the Always-NP Grammar as well because the grammar gets complex, but it only drops to a lesser degree, from -6.5051 to -9.8061. It is also noteworthy that there is not much difference between the two grammars in their likelihoods. This seems to be because corpus extension costs both grammar types, but it does so in different manners: while the Always-NP Grammar type suffers from having to generate trees with more NPs embedded, the Sometimes-NP Grammar type suffers from having to choose from six different rules every time it expands an S.

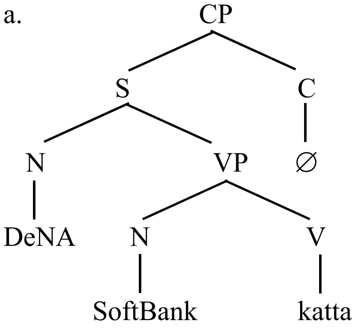
This way, the system allows us to see how grammars of one type describe the data more efficiently than grammars of the other type.

3. Comparing Two Grammar Types for Japanese

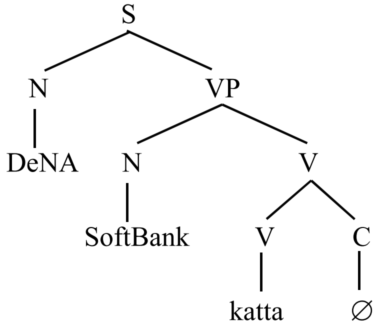
3.1. The First Try

We are now prepared to apply PTR's evaluation metric to the two grammar types introduced in Section 1. The simple declarative sentence *DeNA-ga SoftBank-ni katta* 'The DeNA BayStars beat the SoftBank Hawks' is diagramed in a CH Grammar and HH Grammar, as in (13a, b), respectively. (To simplify the calculations of prior and likelihood probabilities, we focus on grammars where N is not dominated by NP.)

(13) a.

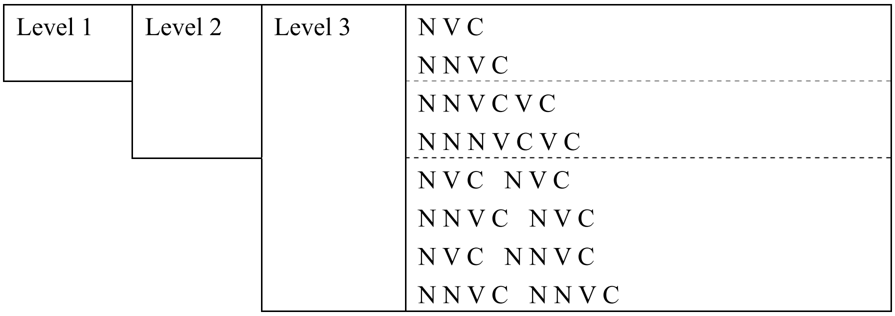


b.



3.1.1. The Corpus

We use a toy corpus shown in Figure 1 as our language. Following PTR, we divide the main corpus into different sub-corpora in order to see if the likelihood of a grammar varies depending on the complexity of the corpus used. We create three corpus levels. Level 1 contains most basic patterns, simple transitive and intransitive sentences. Multi-clause sentences are added at Level 2. They have complement clauses headed by intransitive and transitive verbs. At the third level, the patterns we consider as containing adjunct clauses are added. Both the main clauses and the adjunct clauses are varied in terms of their transitivity. See Table 6 for illustrations of these sentence patterns with Japanese examples.



Level 1	Level 2	Level 3	N V C N N V C
			N N V C V C N N N V C V C
			N V C N V C N N V C N V C N V C N N V C N N V C N N V C

Figure 1. Three-level corpus and the sentence patterns that they comprise.

Table 6. Strings of symbols in the corpus and possible natural language examples of them.

String	Example
N V C	*DeNA-ga katta.* ‘DeNA won.’
N N V C	*DeNA-ga SoftBank-ni katta.* ‘DeNA beat SoftBank.’
N N V C V C	*Hiroshi-wa DeNA-ga katta-to itta.* ‘Hiroshi said that DeNA had won.’
N N N V C V C	*Hiroshi-wa DeNA-ga SoftBank-ni katta-to itta.* ‘Hiroshi said that DeNA had beaten SoftBank.’
N V C N V C	*Hiroshi-ga kaetta-node DeNA-ga katta.* ‘Because Hiroshi left, DeNA won.’
N N V C N V C	*Hiroshi-ga siai-o mita-node DeNA-ga katta.* ‘Because Hiroshi watched the game, DeNA won.’
N V C N N V C	*Hiroshi-ga kaetta-node DeNA-ga SoftBank-ni katta.* ‘Because Hiroshi left, DeNA beat SoftBank.’
N N V C N N V C	*Hiroshi-ga siai-o mita-node DeNA-ga SoftBank-ni katta.* ‘Because Hiroshi watched the game, DeNA beat SoftBank.’

The corpus is not realistic in that it contains too few patterns and that the levels do not reflect the way the complexity of the learner’s input changes during the course of acquisition.

The sentence patterns in our corpus, however, are found in real child-directed speech. The utterances in (14) are all from the Arika corpus of CHILDES (Oshima-Takane et al. 1998, Nisisawa and Miyata 2009). It should be noted that here again, we abstract away from many grammatical features of Japanese (e.g., argument drop, right dislocation).

(14) a. N V C

nanka	ugoita.
something	moved

‘Something moved.’ (ArikaM/40926.cha)

b. N N V C

Arichan kyoo no	gyoza wa	joozu ni	tukutte ne.
Arichan today GEN	dumpling TOP	well COP	make SFP

‘Ari-chan, I hope you’ll make good dumplings today.’ (ArikaM/50004.cha)

c. N N V C V C

kore	zembu	aru	ka doo ka	mite	miyoo.
this	all	exist	whether	see	let’s.see

‘Let’s see if we have all of this here.’ (ArikaM/30900.cha)

d. N N N V C V C

kore	nani	nonde	(i)ru	to	omou ,,	kono hito.
this	what	drink	PROG	that	think	this person

‘What do you think he is drinking here.’ (ArikaM/40610.cha)

e. N N V C N N V C

ima hitotsu no	omocha o	dashitara	hitotsu no	omocha wa	shimau nda.
now one.CL GEN	toy ACC	take.if	one.CL GEN	toy TOP	put.way SFP

‘If you take a toy out, you should put way one, okay?’ (ArikaM/39421.cha)

These child-directed utterances suggest that the corpus we discuss is not entirely hypothetical.

3.1.2. The Grammars

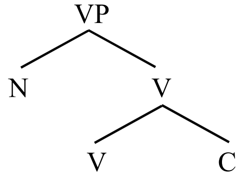
To generate the subsets of the corpus with the minimum number of rules, we posit three grammars for each grammar type, as shown in Figure 2. At Level 1, the grammars are designed to cover the sentences contained in the Level 1 sub-corpus. A Level 2, they are designed to cover the sentences with complement clauses as well because the sub-corpus contains them. Finally, at Level 3, the two grammars are extended so that they can generate up to the complex sentences containing adverbial clauses adjoined in sentence initial position.

To calculate the priors for these six grammars, we need to decide the number of non-terminals, terminals, and rewriting rules. The task is trivial for CH Grammars, but it is not for HH Grammars. This is so because in the latter grammars, the symbol V is used as a terminal symbol as well as a non-terminal symbol. (15) is the structure of a transitive VP under HH Grammars.

			Comp-Head Grammars	Head-Head Grammars
Level 1	Level 2	Level 3	CP → S C S → N VP VP → V | N V	S → N VP VP → V | N V V → V C
			VP → CP V	VP → S V
			S → CP S	S → S S

Figure 2. Two grammars at Levels 1–3 and the rules they comprise. The higher the level is, the more rules each grammar type has.

(15)



Rule	Probability
S → N VP	1
VP → V	0.5
VP → N V	0.5
V → V C	0.5
V: *Do nothing*	0.5

Within the standard conception of context-free grammars (see, e.g., Larson 2010, Carnie 2010), we take this situation to mean that V is ambiguous between being a terminal and being a non-terminal under HH Grammars. Thus, in the Level-1 HH Grammar for example, the grammar's non-terminal symbols include S, VP, and V, whereas its terminal symbols include N, V, and C. This leads to the state of affairs where for each corpus, the CH and HH Grammars do not differ in terms of the numbers of the (non)terminal symbols and production rules, as shown in Table 7.

Table 7. The numbers of the symbols and rules of the CH and HH grammar types at Levels 1–3. At each level, the two grammars do not differ in terms the numbers of non-terminals, terminals and rewriting rules.

	CH Grammar		HH Grammar	
Level 1	# nonterminals	3	# nonterminals	3
	# terminals	3	# terminals	3
	# rules	4	# rules	4
Level 2	# nonterminals	3	# nonterminals	3
	# terminals	3	# terminals	3
	# rules	5	# rules	5

Level 3	# nonterminals	3	# nonterminals	3
	# terminals	3	# terminals	3
	# rules	6	# rules	6

Accordingly, the production rules are assigned production probabilities the way shown in Table 8.

Table 8. Rules of the CH and HH grammars and their probabilities at Levels 1–3.

	CH Grammar		HH Grammar	
	Rule	Prob.	Rule	Prob.
Level 1	CP → S C	1	S → N VP	1
	S → N VP	1	VP → V	1/2
	VP → V	1/2	VP → N V	1/2
	VP → N V	1/2	V → V C	1/2
			V: *Do nothing*	1/2
Level 2	CP → S C	1	S → N VP	1
	S → N VP	1	VP → V	1/3
	VP → V	1/3	VP → N V	1/3
	VP → N V	1/3	VP → S V	1/3
	VP → CP V	1/3	V → V C	1/2
			V: *Do nothing*	1/2
Level 3	CP → S C	1	S → N VP	1/2
	S → N VP	1/2	S → S S	1/2
	S → CP S	1/2	VP → V	1/3
	VP → V	1/3	VP → N V	1/3
	VP → N V	1/3	VP → S V	1/3
	VP → CP V	1/3	V → V C	1/2
			V: *Do nothing*	1/2

3.1.3. Results and Discussion

Now let us calculate the prior, likelihood, and posterior probabilities for each grammar. The results are shown in Table 9. At each level, the two grammar types happen to have the same prior probability. What makes a difference is their likelihood probabilities. At each level, the CH Grammar scores better than its HH counterpart.

Table 9. Log prior, likelihood, and posterior probabilities for CH and HH Grammars. The higher probabilities are highlighted.

	CH Grammar			HH Grammar		
	Prior	Likelihood	Posterior	Prior	Likelihood	Posterior
Level 1	-9.6614	-0.6020	-10.2635	-9.6614	-1.8061	-11.4676
Level 2	-12.120	-2.8627	-14.9835	-12.120	-6.4750	-18.5959
Level 3	-14.580	-12.0982	-26.6785	-14.580	-20.527	-35.1073

To see why the results have turned out this way, consider how the two grammar types assign a probability to the string *N N V C V C*, which corresponds to *Hiroshi-wa DeNA-ga kattta-to itta* ‘Hiroshi said that the DeNA had won’. As shown in (16), the CH Grammar assigns 1/9 to the string of symbols whereas the HH Grammars assigns 1/36 to the same string. Crucial is that in HH Grammars, every time a V node is encountered, a choice has to be made as to whether to expand *V* or not. If the V node in question is a nonterminal, it is rewritten to V C; if it is a terminal, no more rules apply.

(16) a. *CH Grammar*
$[_{CP}$ $[_{S}$ N $[_{VP}$ $[_{CP}$ $[_{S}$ N $[_{VP}$ V]] C] V]] C]
1 1 0.33 1 1 0.33

b. *HH Grammar*
$[_{S}$ N $[_{VP}$ $[_{S}$ N $[_{VP}$ $[_{V}$ V C]]] $[_{V}$ V C]]]
1 0.33 1 0.33 $\underline{0.5}$ $\underline{0.5}$

Generally speaking, if a sentence has more verbs, it lowers the probability of the sentence to a greater degree in a HH Grammar than in a CH Grammar. For this reason, the CH Grammars score better in likelihood probability than the HH Grammars at all the three levels.

3.2. Towards a More Linguistically-Informed Comparison of the Two Grammar Types

3.2.1. BPS-Theoretic Considerations

Does the finding reported in the previous subsection lead us to conclude that CH grammars are superior to HH Grammars? We argue that it does not necessarily. Recall that CH Grammars are more highly valued in likelihood probability, not prior probability. Recall also that when a HH Grammar’s likelihood is calculated, every V node is assigned probability 0.5 because there is a choice as to whether to expand the node or terminate the rewriting process there. We argue that the reason for CH Grammars being more highly valued is essentially because the framework of context-free grammars that underlie the PTR model forces us to posit two kinds of V under HH Grammars. Thus, the significance of the results seen above depends how substantial the terminal-nonterminal distinction is in linguistic theory.

In fact, Bare Phrase Structure (Chomsky 1995: Ch. 4), where bar-level notations are eliminated, suggests that the terminal-nonterminal distinction might not be conceptually necessary. Under the latter theory, CH Grammars would parse a transitive sentence as in (17a) on the one hand while HH Grammars as in (17b) on the other. Crucially, the grammatical asymmetry between the two grammar types seems to vanish. Now both grammars involve terminal-nonterminal ambiguities.

How does this change affect the results? Table 10 lists the rules of the GH and HH Grammars at the three levels and their production probabilities. The final results are shown in Table 11.

(17)

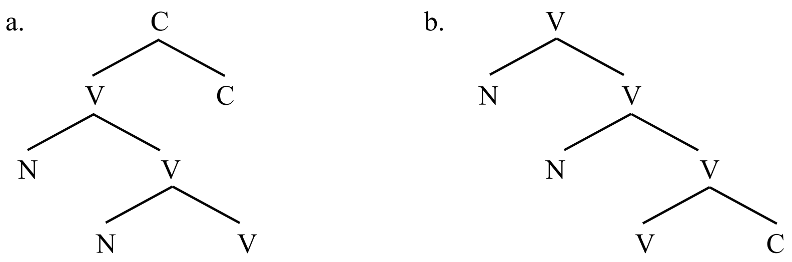


Table 10. Rules of the BPS versions of CH and HH Grammars and their probabilities at Levels 1–3.

	BPS CH Grammar		BPS HH Grammar	
	Rule	Prob.	Rule	Prob.
Level 1	C → V C	1/2	V → N V	1/3
	C: *Do nothing*	1/2	V → V C	1/3
	V → N V	1/2	V: *Do nothing*	1/3
	V: *Do nothing*	1/2		
Levels 2 and 3	C → V C	1/2	V → N V	1/4
	C: *Do nothing*	1/2	V → V V	1/4
	V → N V	1/3	V → V C	1/4
	V → C V	1/3	V: *Do nothing*	1/4
	V: *Do nothing*	1/3		

Table 11. Log prior, likelihood, and posterior probabilities for 'bare phrase structure' CH and HH Grammars. At each level, the higher prior, likelihood and posterior probabilities are highlighted.

	BPS CH Grammars			BPS HH Grammars		
	Prior	Likelihood	Posterior	Prior	Likelihood	Posterior
Level 1	-5.0103	-2.7093	-7.7196	-4.8676	-3.3398	-8.2075
Level 2	-7.5051	-11.2463	-18.7514	-6.6226	-13.2453	-19.8679
Level 3	-7.5051	-27.5136	-35.0188	-6.6226	-32.5112	-39.1338

The results reveal that at all the corpus levels, the CH Grammars are preferred over the HH Grammars. The reason seems to be that production probabilities in HH Grammars are generally lower than those in CH Grammars, as can be observed in Table 10. This generally causes the HH Grammars' likelihood probabilities lower.

3.2.2. Further Discussion

What can we conclude from the evaluations of the CH and HH Grammars in the BPS setting? We are now more certain than before that CH Grammars are slightly but constantly preferred over their HH counterparts. While the arguably unsubstantial asymmetry between the grammar types is eliminated (i.e., both grammar types now involve the terminal/nonterminal ambiguity in question), the CH Grammars score better than the HH Grammars, even if the difference may be small.

These results with the second model, however, are hard to interpret, we have to say. First, if BPS Theory is what we should pursue here, the grammatical framework underlying the PTR evaluation metric is at odds with that. When the complexity of a particular grammar is evaluated, we should not look at the number of terminals and nonterminals to measure its value if we think that these notions are groundless.

Furthermore, it should be noted that the BPS grammars discussed above are empirically inadequate in many ways. Among other things, they massively overgenerate ungrammatical strings unlike the more traditional theories that we adopt in the first model. For example, assuming that C and V are considered as start symbols in BPS CH and HH Grammars respectively, both grammars generate the string *V C*, which is considered ungrammatical. The cause of overgeneration of this sort obviously is the lack of subcategorization in context-free grammars. Transitive and intransitive verbs, for instance, require subject arguments. This requirement is encoded by the S-rules in the non-BPS grammars. And it is not possible to use an S-rule in BPS grammars. The situation is not surprising at all. It has been noticed since Stowell (1981) that phrase structure rules like those we consider in the previous section implicitly encode what subcategorization features do, and for this very reason, they are considered redundant. The BPS versions of them have no redundancy in this respect but, as a tradeoff, lose the ability to capture the restrictions on what elements each verb can cooccur with.[5] Problems of the BPS grammars like these make them far from realistic as theories of Japanese syntax.

Thus, one future task would be to reconsider the BPS grammars underlying our evaluation metric in a way that allows us to handle subcategorization. Are there any grammars we can adopt for that purpose? Minimalist Grammars (Stabler 1997, Kobel 2006, Hunter 2010, Graf 2013) are an obvious candidate, although it takes us too far afield to ask at this point what the priors and the likelihoods for Minimalist CH and HH Grammars should look like.[6]

4. Conclusion

We have applied Perfors, Tenenbaum, and Regier's (2006, 2011) model of sentence structure acquisition to comparison of two grammar types for Japanese. Although the results of the present study may be taken as suggesting that Complement-Head Grammars be more highly valued than Head-Head Grammars, we argue that cautious interpretation is required. We hope

---

[5] Other problems include the fact that neither grammar can distinguish complements and adjuncts, though the problem is not inherent to the context-free interpretation of BPS theory (Hornstein and Nunes 2008). That is why no new rules are added to a grammar when the corpus is extended from Level 2 to Level 3.

[6] Incidentally, many analyses that adopt a CH grammar assume head movement (Otani and Whiman 1991; Koizumi 2000; Kishimoto 2008; Funakoshi 2014, 2017, 2020; Hayashi and Fujii 2015, 2016; Sato and Hayashi 2018, among many others). This feature cannot be added to any of the grammars we have considered here but can be added to Minimalist Grammars.

to have shown that learning models of the sort discussed here may be useful for theory selection and that theoretical considerations on the grammars underlying them are crucial in order to obtain meaningful results.

References

Aoyagi, Hiroshi (2006) *Nihongo no Zyosi to Kinoohantyuu* [Japanese Particles and Functional Heads], Hituzi Syobo, Tokyo.

Carnie, Andrew (2010) *Constituent Structure*, Oxford University Press, Oxford, UK.

Chomsky, Noam (1965) *Aspects of Theory of Syntax*, MIT Press, Cambridge, Mass.

Chomsky, Noam (1970) “Remarks on Nominalization.” In *Reading in English Transformational Grammar*, R. Jacobs and Peter Rosenbaum, eds., Ginn, Waltham, 184–221.

Chomsky, Noam (1995) *The Minimalist Program*, MIT Press, Cambridge, Mass.

Chomsky, Noam, and Morris Halle (1968) *Sound Patterns of English*, Harper & Row, New York.

Farmer, Ann (1980) *Modularity in Syntax*, MIT Press, Cambridge, Mass.

Freidin, Robert, and Howard Lasnik (2011) “Some Roots of Minimalism in Generative Grammar.” In *The Oxford Handbook of Linguistic Minimalism*, Cedric Boeckx, ed., Oxford University Press, Oxford, 1–26.

Fujii, Tomohiro (2016) “Hobun no Kozo to Umekomi Hobun no Bunrui (Structure of Complement Clauses and Classification of Embedded Clauses).” In *Nihongo Bunpo Handobukku* (Japanese Grammar Handbook), Keiko Murasugi, Mamoru Saito, Yoichi Miyamoto and Kensuke Takita, eds., Tokyo, Kaitakusha, 2–37.

Fukui, Naoki, and Hiromu Sakai (2003) “The Visibility Guideline for Functional Categories: Verb Raising in Japanese and Related Issues,” *Lingua* 113, 321–375.

Fukui, Naoki, and Yuji Takano (1998) “Symmetry in Syntax: Merge and Demerge,” *Journal of East Asian Linguistics* 7, 27–86.

Funakoshi, Kenshi (2014) *Syntactic Head Movement and Its Consequences*. Ph.D. dissertation, University of Maryland.

Funakoshi, Kenshi (2017) “Verb-Stranding Verb Phrase Ellipsis in Japanese,” *Journal of East Asian Linguistics* 25, 113–142.

Funakoshi, Kenshi (2020) “Verb-Raising and VP-Fronting in Japanese,” *The Linguistic Review* 37, 117–146.

Graf, Thomas (2013) *Local and Transderivational Constraints in Syntax and Semantics*, Ph.D. dissertation, University of California, Los Angeles.

Han, Chung-hye, Jeffrey Lidz, and Julien Musolino (2007) “V-Raising and Grammar Competition in Korean: Evidence from Negation and Quantifier Scope,” *Linguistic Inquiry* 38, 1–47.

Han, Chung-hye, Jeffrey Lidz, and Julien Musolino (2016) “Endogenous Sources of Variation in Language Acquisition,” *Proceedings of the National Academy of Sciences* 113, 942–947.

Hayashi, Shintaro, and Tomohiro Fujii (2015) “String Vacuous Verb Raising: The Case of V-*te* in Japanese,” *Gengo Kenkyu* 147, 31–55.

Hayashi, Shintaro, and Tomohiro Fujii (2016) “Predicate Ellipsis, Null Adjuncts, and Predicate Raising in Japanese,” *Proceedings of CLS 51,* 225–239.

Hornstein, Norbert, and Jairo Nunes (2008) “Adjunction, Labeling, and Bare Phrase Structure,” *Biolinguistics* 2, 57–86.

Hunter, Tim (2010) *Syntactic Effects of Conjunctivist Semantics: Unifying Movement and Adjunction*, John Benjamins, Amsterdam/Pittsburg.

Kayne, Richard (1994) *The Antisymmetry in Syntax*, MIT Press, Cambridge, Mass.

Kishimoto, Hideki (2008) "On Verb Raising." In *The Oxford Handbook of Japanese Linguistics,* Shigeru Miyagawa and Mamoru Saito, eds., Oxford University Press, Oxford, 107–140.

Kitagawa, Yoshihisa (1986) *Subject in Japanese and English*, Ph.D. dissertation, University of Massachusetts, Amherst.

Kobele, Gregory M. (2006) *Generating Copies: An Investigation into Structural Identity in Language and Grammar*, Ph.D. dissertation, University of California, Los Angeles.

Koizumi, Masatoshi (2000) "String Vacuous Overt Verb Raising," *Journal of East Asian Linguistics* 9, 227–285.

Koopman, Hilda (2005) "Korean (and Japanese) Morphology from a Syntactic Perspective," *Linguistic Inquiry* 36, 601–633.

Kornai, András, and Geoffrey K. Pullum (1990) "The X-bar Theory of Phrase Structure," *Language* 66, 24–50.

Kuroda, S.-Y. (1965) *Generative Grammatical Studies of the Japanese Language*, Ph.D. dissertation, MIT.

Kuroda, S.-Y. (2003) "Complex Predicates and Predicate Raising," *Lingua* 113: 447–480.

Larson, Richard (2010) *Grammar as Science*, MIT Press, Cambridge, Mass.

Miyagawa, Shigeru (1980) *Complex Verbs and the Lexicon*, Ph.D. dissertation, University of Arizona.

Miyagawa, Shigeru (1999) "Causatives." In *The Handbook of Japanese Linguistics*, Natsuko Tsujimura, ed., Blackwell, 236–268.

Murasugi, Keiko (2000) "Japanese Complex Noun Phrases and the Antisymmetry Theory." In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, Roger Martin, David Michals, and Juan Uriagereka, eds., MIT Press, Cambridge, Mass., 211–233.

Nisisawa, Hiro Yuki, and Susanne Miyata (2009) *MiiPro-Arika Corpus*, TalkBank, Pittsburg.

Otani, Kazuyo, and John Whitman (1991) "V-Raising and VP-Ellipsis," *Linguistic Inquiry* 22, 345–358.

Oshima-Takane, Yuriko, and Brian MacWhinney, Hidetoshi Sirai, Susanne Miyata, Norio Naka (1998) *CHILDES Manual for Japanese*, *2nd Edition*. The JCHAT project, Chukyo University.

Pearl, Lisa (2017) "Evaluation, Use, and Refinement of Knowledge Representations Through Acquisition Modeling," *Language Acquisition* 24, 126–147.

Pearl, Lisa (2020) "Poverty of the Stimulus Without Tears," unpublished manuscript, University of California, Irvine.

Perfors, Andrew, Joshua B. Tenenbaum, and Terry Regier (2006) "Poverty of the Stimulus? A Rational Approach." In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, Ron Sun and Naomi Miyake, eds., New York & London, Psychology Press, 663–668.

Perfors, Andrew, Joshua B. Tenenbaum, and Terry Regier (2011) "The Learnability of Abstract Syntactic Principles," *Cognition*, 118: 306–338.

Perfors, Andrew, Joshua B. Tenenbaum, Thomas L. Griffiths, and Fei Xu (2011) "A Tutorial Introduction to Bayesian Models of Cognitive Development," *Cognition* 120, 302-321.

Saito, Mamoru (2012) "Case Checking/Valuation in Japanese: Move, Agree or Merge?" *Nanzan Linguistics* 8, 109–127.

Saito, Mamoru, and Naoki Fukui (1998) "Order in Phrase Structure and Movement," *Linguistic Inquiry* 29, 439–474.

Saito, Mamoru, and Hitoro Hoshi (1998) "Control in Complex Predicates," Paper presented at the Tsukuba Workshops on Complex Predicates, 1998.

Saito, Mamoru, and Hiroto Hoshi (2000) "The Japanese Light Verb Construction and the Minimalist Program." In *Step by Step: Essays on Minimalist Syntax in Honor of Howard Lasnik*, Roger Martin, David Michaels and Juan Uriagereka, eds., MIT Press, Cambridge, Mass., 261–295.

Sato, Yosuke, and Shintaro Hayashi (2018) "String-Vacuous Head Movement in Japanese: New Evidence from Verb-Echo Answers," *Syntax* 21, 72–90.

Sells, Peter (1995) "Korean and Japanese Morphology from a Lexical Perspective," *Linguistic Inquiry* 26, 277–325.

Shibata, Yoshiyuki (2015) "Negative Structure and Object Movement in Japanese," *Journal of East Asian Linguistics* 24, 217–269.

Shimada, Junri (2007) "Head Movement, Binding Theory, and Phrase Structure," unpublished manuscript, MIT.

Stabler, Edward (1997) "Derivational Minimalism," In Christian Retoré, ed., *Logical Aspects of Computational Linguistics*, *Volume 1328 of Lecture Notes in Computer Science,* Berlin, Springer-Verlag, 68–95.

Stowell, Tim (1981) *Origins of Phrase Structure*, Ph.D. dissertation, MIT.

Takita, Kensuke (2009) "If Chinese is Head-Initial, Japanese Cannot Be," *Journal of East Asian Linguistics* 18, 41–61.

Takita, Kensuke (2013) "V-Stranding Sluicing and Its Implications." In *Proceedings of NELS 41*, Nicholas LaCara, Yelena Fainleib and Yangsook Park, eds., GLSA, Amherst, Mass., 247–260.

Tokizaki, Hisao, and Yasutomo Kuwana (2013) "A Stress-Based Theory of Disharmonic Word Orders." In *Theoretical Approaches to Disharmonic Word Orders*, Theresa Biberauer and Michelle Sheehan, eds., Oxford University Press, Oxford, 190–215.

Tonoike, Shigeo (1995) "Japanese as an OSV Language." In Shosuke Haraguchi and Michiko Funaki, eds., *Minimalism and Linguistic Theory*, Tokyo, Hituzi Shobo, 105–141.

Whitman, John (1998) "Gojun to Togokozo (Word Order and Syntactic Structure)." In *Kaku to Gojun to Togokozo* (Case, Word Order, and Syntactic Structure) by Koichi Takezawa and John Whitman, Tokyo, Kenkyusha, 104–195.

Yang, Charles (2017) "Rage Against the Machine: Evaluation Metrics in the 21st Century," *Language Acquisition: A Journal of Developmental Linguistics* 24, 100–125.