Maximizing our Chances of Learning from Errors in Language Faculty Science: Suggestions and Illustration*  
Hajime Hoji  
University of Southern California

Abstract

This paper is concerned with how hypotheses about the language faculty can be made testable. Adopting the general model of the Computational System in Chomsky 1993, a model of judgment-making suggested in Ueyama 2010, and a research heuristic "Maximize our chances of learning from errors," the paper suggests how we can try to identify what is likely a reflection of properties of the Computational System hypothesized at the center of the language faculty. Testability pursued here is in terms of "point-value predictions" rather than predictions about a difference or a tendency. The paper argues, as a consequence of the three starting assumptions, that we can obtain categorical judgments, but only if we recognize a fundamental asymmetry between a *Schema-based prediction and an *okSchema-based prediction. The paper provides some illustration by applying the suggested method to lexical hypotheses in Japanese, by making reference to results of on-line experiments. The paper also provides an illustration that it is indeed possible to obtain categorical judgments from informants, contrary to the generally accepted view in the field that evidence for or against hypotheses about language (or the language faculty) can be assessed only in the terms of statistically significant contrasts, as in the general tradition of social, behavioral and life sciences.

1. Introduction

Linguistics is often said to be a scientific study of language, and generative grammar a scientific study of the language faculty. While some people might consider the mention of "scientific" rhetorical, others maintain that it is not. To the extent that we indeed wish to be engaged in a scientific research program that is concerned with the properties of the language faculty, it is worth considering how our hypotheses about the language faculty can be made testable and how we can aspire to make our research program "scientific."

In the generative tradition, the main goal of our research is understood to be a discovery of the properties of the Computational System, hypothesized to be at the center of the language faculty. It is furthermore assumed that a major source of evidence for or against our hypotheses is informant judgments. Despite a wide acceptance of this assumption, however, the field has so far failed to seriously consider in what way informant judgments can be revealing about the properties of the Computational System, let alone come up with an answer that the majority of the field can agree upon. Chomsky 1986: 36, for example, remarks, "In general, informant judgments do not reflect the structure of the language directly; judgments of acceptability, for example, may fail to provide direct evidence as to grammatical status because of the intrusion of numerous other factors," and it has remained unclear

* [To be removed in the version to be submitted to Language, and what is stated below will be revised in the final version.] I am indebted to the late Yuki Kuroda for his encouragement and friendship over the years. I hope that the comments and concerns he conveyed to me in February of 2009 about the content, and especially about the presentation, of the first few chapters of a draft of Hoji 2009 have made this paper better than what it would have been without them, although he might say, with his usual grin, that he is not sure about that. I would also like to thank Yasuo Deguchi, Jean Gerard, Caroline Scherzer, Jennifer Smith for their comments on (an) earlier version(s) of the paper. Detailed comments by Teru Fukaya, Emi Mukai and Kiyoko Kataoka on several draft versions have resulted in much improvement. I have also benefitted from comments by Yasuaki Abe, Chris Kennedy and Hiroki Narita. Discussions with Ayumi Ueyama and Yuki Takubo over the years have helped me form the views presented here. None of the aforementioned is responsible for the remaining errors and shortcomings of the paper.
exactly what should qualify as evidence for or against our hypotheses about the language faculty, how we can identify such evidence, and hence how we can put our hypotheses to rigorous empirical test.\(^1\)

The main thesis of this paper is that we can pursue language faculty science as an "exact science," in which we can extract \textit{reproducible phenomena} from informants' introspective judgments and analyze them \textit{quantitatively}. This seems to go directly against the common perception that informants' (including researchers' own) introspective judgments are not reliable source of data for or against our hypotheses about language faculty because of across-informant and within-informant variation and fluctuation. One can view the main aim of the paper as providing a basic tool in language faculty science in response to Anderson's (2000: 812) "whimper" that "we need better tools of investigation" in language faculty science.

In the seventh lecture of his 1964 Messenger Lectures at Cornell University "Seeking New Laws," Richard Feynman states:

\begin{quote}
In general, we look for a new law by the following process. First we guess it. Then we compute the consequences of the guess to see what would be implied if this law that we guessed is right. Then we compare the result of the computation to nature, with experiment or experience, compare it directly with observation, to see if it works. If it disagrees with experiment, it is wrong. In that simple statement is the key to science. It does not make any difference how beautiful your guess is. It does not make any difference how smart you are, who made the guess, or what his name is—if it disagrees with the experiment, it is wrong. That's all there is to it. (Feynman 1965/94: 150)
\end{quote}

Feynman continues by adding the following "obvious remarks":\(^2\)

\begin{quote}
It is true that one has to check a little to make sure that it is wrong, because whoever did the experiment may have reported incorrectly, or there may have been some feature in the experiment that was not noticed, some dirt or something; or the man who computed the consequences, even though it may have been the one who made the guesses, could have made some mistake in the analysis. These are obvious remarks, so when I say if it disagrees with experiment it is wrong, I mean after the experiment has been checked, the calculations have been checked, and the thing has been rubbed back and forth a few times to make sure that the consequences are logical consequences from the guess, and that in fact it disagrees with a very carefully checked experiment. (Feynman 1965/94: 150-1)
\end{quote}

This paper sketches how the above-mentioned general scientific method, schematized in (1), can be applied to research concerned with the properties of the language faculty.\(^3\)

\footnotesize

1 Since the inception of generative grammar in the 1950s, there has been concern about in what way our hypotheses with regard to the Computational System of the language faculty can be put to (rigorous) empirical test. Chomsky's remarks on the relation between acceptability and grammaticality, which are frequently found in his writings until about the mid 1970s, can be understood to be a manifestation of his own concern. A number of publications in the late 1960s and in the 1970s addressed the issue (e.g., many of the articles included in Cohen 1974, Cohen and Wirth 1975, and Eckman 1977); see Schütze 1996 for an extensive literature review. Concerns seem to have resurfaced in recent years; cf. the 2005 special issue of \textit{Lingua} on "Data in Theoretical Linguistics," the 2007 special issue of \textit{Theoretical Linguistics} volume on "Data in generative grammar," Devitt 2006, and Fitzgerald 2010, among many others.

2 The "obvious remarks" should not be taken as reducing the significance of "the key to science" in the first quote. The point intended in the "obvious remarks" is not that we should not concern ourselves with empirical details and the testability of our hypotheses. That is given. On the contrary, the point of the "obvious remarks" must be about the importance of empirical (as well as theoretical) rigor. The point seems to be either missed or misrepresented in Boeckx 2006, judging from the way Feynman's remarks are cited there (p. 89, footnote 21, for example); see section 6.3 below for related discussion. Similar remarks seem to apply to the way Lakatos' work is cited in Boeckx 2006, as pointed out in Kuroda 2008: footnote 3.

3 What will be discussed below is more concerned with the "Compare with Experiment" part than with the "Guess" and
1. The general scientific method (i.e., the hypothetico-deductive method):

   Guess — Computing Consequences — Compare with Experiment

   One may object that physics may not be the right field for us to turn to. After all, it seems commonly understood that in fields other than physics (and those closely related to it), predictions are about differences and/or tendencies, not about point-values; cf. Meehl 1967: 264 and Barnard et al. 2007, for example. As noted, the main thesis of this paper is that we can make point-value predictions in language faculty science and expect them to be supported by experimental results.

   Section 2 addresses methodological issues and advances the essentials of the proposal for testing our hypotheses about properties of the language faculty. Section 3 provides an illustration of the proposal, discussing so-called local reciprocal anaphor in Japanese. Section 4 offers a summary of the proposed methodology, making reference to two research heuristics. Further illustration is provided in section 5 although the discussion is rather abbreviated due to space considerations. Some implications are discussed in Section 6. In Appendix, I provide the examples and the results of experiments dealing with so-called local reflexive anaphors in Japanese. The language dealt with is Japanese. It is hoped that similar illustration will in the future be made dealing with other languages, in line with what is suggested in section 6.2.

2. Methodological preliminaries

2.1. The goal of generative grammar and the computational system

   In what follows, I use generative grammar to refer to research concerned with the properties of the language faculty, and more in particular with those of the Computational System as it is hypothesized to be at the center of the language faculty and use the adjective generative accordingly. I also assume that a major source of evidence for or against our hypotheses concerning the Computational System is informant judgments, as explicitly stated by N. Chomsky in Third Texas Conference on Problems of Linguistic Analysis in English May 9-12, 1958, published in 1962 by the University of Texas.

   "Compute Consequences" parts. A fuller illustration of what will be suggested below would include more in-depth discussion of the latter.

4 A more satisfactory illustration would require articulation of various issues, including how the various universal as well as language-particular hypotheses in question are formally stated, how the predictions are deduced from them, whether and how each of the testable hypotheses in question has been empirically supported in line with the method of hypothesis testing to be proposed below, a fuller discussion of which cannot be provided here for space reasons.

5 The testing of a universal hypothesis can be done only in the context of a particular language, and designing effective and reliable experiments requires a great deal of understanding of various aspects of the language in question. The choice of Japanese for the purpose of the illustration of the proposed methodology here is similar to the reason why Culicover and Jackendoff 2005's "discussion is mainly focused on English." Culicover and Jackendoff 2005: xiv-xv remark, "This [i.e., the fact that their discussion is mainly focused on English, HH] does not mean that we think linguistics can be studied in the context of English alone, only that we think others can do other languages better."

6 This is perhaps too narrow a characterization of generative grammar; see Culicover and Jackendoff 2005: chapter 1, for example, for remarks on a wide spectrum of research orientations and practices under the name of generative grammar. Furthermore, given the actual practice in the field over the years, one might even object to equating generative grammar with language faculty science. In this paper, we shall not be concerned with the terminological issues and I will use generative (grammar) in the way just noted in the text.

7 Chomsky's remarks in Third Texas Conference on Problems of Linguistic Analysis in English seem to point directly to what he had in mind at least around 1958, in my view more directly than what we typically find in his writings in the 1950s and 1960s and the subsequent years. One of many such remarks by Chomsky in that volume is reproduced here (p. 168 of the 1958 volume); see also Chomsky 1986: 36-37. Hill: Linguistic intuition is itself a system, almost a complete grammar. If it is good enough, why bother with any other grammar?
Minimally, the language faculty must relate 'sounds' (and signs in a sign language) and 'meanings'. A fundamental hypothesis in generative grammar is the existence of the Computational System at the center of the language faculty. Since Chomsky 1993, it is generally understood in generative research that the Computational System takes as its input a set of items in the mental Lexicon of speakers of a language and yields as its output a pair of mental representations—one underlying sounds/signs and the other 'meanings'. Following the common practice in the generative tradition since the mid-1970s, let us call the former a PF (representation) and the latter an LF (representation). The model of the Computational System (CS) as suggested in Chomsky 1993 can be schematized as in (2).

(2) The Model of the Computational System:

\[
\text{Numeration } \mu \Rightarrow \text{CS} \Rightarrow \text{LF(}\mu\text{)} \leftarrow \text{PF(}\mu\text{)}
\]

**PF(\mu)**: a set of items taken from the mental Lexicon

**LF(\mu)**: an LF representation based on \( \mu \)

**PF(\mu)**: a PF representation based on \( \mu \)

The PF and the LF representations in (2) are meant to be abstract representations that underlie a sequence of sounds/signs and its interpretation, respectively. Specific implementations of the leading idea behind (2), as they have been suggested and pursued in works subsequent to Chomsky 1993, are inconsequential to the present discussion as far as I can tell; they would be only if they would contribute to yielding testable predictions distinct from what will be discussed below. Our hypotheses about the Computational System are thus meant to be about what underlies the language users' intuitions about the relation between sounds/signs and 'meanings' as reflections of properties of the Computational System. The main goal of generative grammar can be understood as demonstrating the existence of the Computational System by discovering its properties.8

2.2. The model of judgment-making

As noted, the language faculty must relate sounds/signs and 'meanings'. By adopting the thesis that informant judgments are a primary source of evidence for or against hypotheses concerning the Computational System, we are committing ourselves to the view that informant judgments are, or at least can be, revealing about properties of the Computational System. While it may not be obvious how, it seems reasonable to assume that the Computational System is 'made use of' during the act of judgment-making. For, otherwise, it would not be clear how informant judgments could be taken as evidence for or against our hypotheses about the Computational System. We can schematically express this as in (3).

(3) Embedding the Computational System in the model of judgment-making:

\[\gamma(a, b)\]

---

8 Construed in this way, it is not language as an 'external object' but the language faculty that constitutes the object of inquiry in generative grammar, as stated explicitly in Chomsky 1965: chapter 1.
a. $\gamma(a, b)$: an intuition that two linguistic expressions $a$ and $b$ are related in a particular manner\(^9\)

b. $\alpha$: presented sentence

c. $\beta$: the informant judgment on the acceptability of $\alpha$ under $\gamma(a, b)$

The boxed part in (3) is the Computational System; see (2). The informant is presented sentence $\alpha$ and asked whether it is acceptable, or how acceptable it is, under a particular interpretation $\gamma(a, b)$ involving two linguistic expressions $a$ and $b$. As noted above, insofar as informant judgments are assumed to be revealing about properties of the Computational System, the Computational System must be involved in the act of judgment-making by the informant. Given that a numeration is input to the Computational System, it thus seems reasonable to hypothesize that, when making his/her judgment, the informant comes up with a numeration $\mu$ and compares (i) the two output representations based on $\mu$ with (ii) the 'sound' (i.e., the presented sentence $\alpha$) and the relevant 'meaning' under discussion (i.e., the interpretation $\gamma(a, b)$).

The following model of judgment-making by informants presents itself.\(^{10}\)

The Model of Judgment-Making by the Informant on the acceptability of sentence $\alpha$ with interpretation $\gamma(a, b)$\(^{11}\) (based on A. Ueyama’s proposal):

\[\begin{array}{c}
\alpha \Rightarrow \mu \\
\Rightarrow \text{CS} \Rightarrow \text{LF}(\mu) \Rightarrow \text{SR}(\mu) \\
\Rightarrow \text{PF}(\mu) \\
\end{array}\]

\[\begin{array}{c}
\gamma(a, b) \\
\Rightarrow \beta \\
\end{array}\]

a. $\alpha$: presented sentence

b. $\mu$: numeration

c. $\gamma(a, b)$: the interpretation intended to be included in the 'meaning' of $\alpha$ involving expressions $a$ and $b$

d. LF($\mu$): the LF representation that obtains on the basis of $\mu$

e. SR($\mu$): the information that obtains on the basis of LF($\mu$)

\[\begin{array}{c}
Pf(\mu) \\
\end{array}\]

\[\begin{array}{c}
\alpha \\
\Rightarrow \text{Extract} \Rightarrow \mu \\
\Rightarrow \text{CS} \\
\Rightarrow \text{PF}(\mu) \\
\end{array}\]

\[\begin{array}{c}
\gamma(a, b) \\
\Rightarrow \beta \\
\end{array}\]

9 Among the examples of "an intuition that two linguistic expressions $a$ and $b$ are related in a particular manner" are so-called anaphoric dependency, dependency of so-called variable binding and so-called scope dependency. If we represent the relevant relations as $\gamma$, $\gamma'$, and $\gamma''$, they can be expressed as $\gamma$(John, himself) in John praised himself, $\gamma$(everyone, his) in everyone praised his family, and $\gamma''$(everyone, someone) in everyone praised someone. See footnote 11 below.

10 The model in (4), which is adapted from the proposal in a series of works by Ayumi Ueyama, including Ueyama 2010, can be understood as characterizing a specialized instance of the model of comprehension. It may be well to emphasize that the act of judgment-making, more often than not, requires that informants do something that is not involved in ordinary language use. As I hope will be made clear in the ensuing discussion, such idealization is necessary in extracting 'information' pertaining to the properties of the Computational System from informant judgments. It may be an interesting exercise to compare (4) with the model of comprehension discussed in Townsend and Bever 2001 and with the model of judgment-making suggested in Schütze 1996: 175.

11 It would be qualitatively more difficult to maximize our chances of learning from errors if we dealt with simple (un)acceptability without considering $\gamma(a, b)$. I will return to this issue in section 4.2.
g. \( pf(\mu) \): the surface phonetic string that obtains on the basis of \( PF(\mu) \)

h. \( \beta \): the informant judgment on the acceptability of \( \alpha \) under \( \gamma(a, b) \)

The "\( \Rightarrow \)" in (4) indicates that a numeration is input to the Computational System (CS) and its output representations are LF and PF, and that SR and \( pf \) obtain based on LF and PF, respectively. What is intended by "\( \sim \Rightarrow \)" on the other hand, is not an input/output relation, as roughly indicated in (5).

(5) a. Presented Sentence \( \alpha \sim \Rightarrow \) Numeration Extractor: ... contributes to ...

b. Numeration Extractor \( \sim \Rightarrow \) numeration \( \mu \): ... forms ...

c. SR(\( \mu \)) \( \sim \Rightarrow \) Judgment \( \beta \): ... serves as a basis for ...

2.3. Informant judgments and the fundamental asymmetry

Crucial for making testable predictions is a claim—which we may call a bridging statement—that \( \gamma(a, b) \) (see (4b) and footnote 9 and 11) arises only if what corresponds to \( a \) stands in a certain structural relation with what corresponds to \( b \) at LF.\(^{15}\) It seems reasonable to assume that the informant judgment \( \beta \) can be affected by difficulty in parsing and the unnaturalness of the interpretation of the entire sentence in question. Therefore, even if the informant (eventually) "finds" a numeration \( \mu \) that would result in \( pf(\mu) \) non-distinct from \( \alpha \) and SR(\( \mu \)) compatible with the interpretation \( \gamma(a, b) \), that may not necessarily result in the informant reporting that \( \alpha \) is (fully) acceptable under \( \gamma(a, b) \). On the other hand, if the informant fails to come up with such a numeration \( \mu \), the informant's judgment on \( \alpha \) under \( \gamma(a, b) \) should necessarily be "complete unacceptability." For, in that case, the informant fails to "arrive at" SR(\( \mu \)) compatible with the interpretation \( \gamma(a, b) \) presumably because the (hypothesized) structural condition necessary for \( \gamma(a, b) \) is never met in any LF(\( \mu \)) no matter what possible \( \mu \) might be tried. This is the source of the fundamental asymmetry between a *Schema-based prediction and an *okSchema-based prediction (to be introduced in the next subsection) in terms of the significance of their failure.

The failure to understand the asymmetry seems to me to have resulted in a great deal of confusion in the field over the years, including how to deal with, and assess, judgmental variations, fluctuation and disputes—their significance would differ tremendously, depending upon which of the two types of predictions is being addressed. This also seems to me to have contributed to the formation of the common perception that it is not possible to obtain informant judgments of a categorical nature, leading to the view that it is not possible to make "point-value" predictions in research concerned with the language faculty. The asymmetry will play the most crucial conceptual basis of what will be presented in this paper.

2.4. Empirical rigor, "facts," and confirmed schematic asymmetries

The minimal empirical prerequisite for effective pursuit of the discovery of the properties of the language

\(^{12}\) The introduction of \( SR \) and \( pf \) is not crucial for the purpose of the empirical discussion in this paper, and equating \( LF \) and \( PF \) to \( SR \) and \( pf \), respectively, would not affect the ensuring discussion.

\(^{13}\) Numeration formation does not require a presented sentence, as can be seen by the fact that a numeration can be formed in sentence production without a presented sentence. In addition to the presented sentence \( \alpha \) and the interpretation intended to be included in the 'meaning' of \( \alpha \) involving expressions \( a \) and \( b \) (i.e., \( \gamma(a, b) \)), the mental act of numeration formation must also make reference to the mental Lexicon. It must also make reference to the knowledge accumulated through the years of linguistic experiences concerning what Ueyama calls "frequent patterns," i.e., knowledge as to the tendencies of what kind of linguistic expressions go with what kind of a predicate, for example, although that is not indicated in the chart in (4).

\(^{14}\) More precisely, what serves as a basis for the informant judgment \( \beta \) is the compatibility between SR(\( \mu \)) and \( \gamma(a, b) \) or the lack thereof. The non-distinctness between \( pf(\mu) \) and the presented sentence \( \alpha \) also serves as a basis for \( \beta \), but that is not indicated in (4).

\(^{15}\) Predictions here cover both predictions and retrodictions.
faculty is being able to identify informant intuitions that are likely a reflection of properties of the Computational System hypothesized to be at the center of the language faculty. Without being able to do so, neither could we specify the consequences of "our guess" about the Computational System nor could we compare them with the results of a "very carefully checked experiment."

I would like to suggest that what we can regard as a likely reflection of properties of the Computational System is a confirmed schematic asymmetry such that sentences conforming to one type of Schema are always judged to be completely unacceptable under a specified interpretation while those conforming to the other type of Schema, minimally different from the former in terms of the hypothesized formal property, are not necessarily judged to be completely unacceptable. The asymmetry follows from the considerations given in the preceding subsections. Let us refer to the former type of Schema as a *Schema* (which can be read as "star schema") and sentences conforming to it as *Examples* (which can be read as "star examples") and the latter type of Schema as an okSchema and sentences conforming to it as okExamples.

A *Schema*-based prediction is as in (6), and one of the possible formulations of an okSchema-based prediction is as given in (7):

(6) A *Schema*-based prediction:
Informants judge any *Example* conforming to a *Schema* to be completely unacceptable under interpretation $\gamma(a, b)$.

(7) An okSchema-based prediction—version 1: 16
Informants judge okExamples conforming to an okSchema to be acceptable (to varying degrees) under interpretation $\gamma(a, b)$.

There are two crucial points intended by schematic asymmetries. One is that the contrast of significance is not between examples but it is between Schemata. The other is that the contrast must be such that a *Schema*-based prediction has survived a rigorous disconfirmation attempt and is accompanied by the confirmation of the corresponding okSchema-based prediction(s).

While the formulation of a *Schema*-based prediction in (6) is "definitive," so to speak, there is a continuum of formulations for an okSchema-based prediction. Instead of (7), one can adopt (8), for example, which is less stringent than (7) because the existence of just one okExample that is judged to be acceptable would confirm (8).

(8) An okSchema-based prediction—version 2:
Informants judge some okExample conforming to an okSchema to be acceptable (to varying degrees) under interpretation $\gamma(a, b)$.

If we adopt the formulation of an okSchema-based prediction in (7) or (8)—taking the formulation of a *Schema*-based prediction in (6) as 'invariant'—, we can state the fundamental asymmetry as follows: okSchema-based predictions cannot be disconfirmed and they can only be confirmed; *Schema*-based predictions, on the other hand, can be disconfirmed although they cannot be confirmed because it is not possible to consider all the *Examples* that would conform to a *Schema*.

If there is no numeration $\mu$ corresponding to $\alpha$ that would result in (i) LF($\mu$) (hence SR($\mu$)) compatible with $\gamma(a, b)$ and (ii) PF($\mu$) (hence pf($\mu$)) non-distinct from $\alpha$, we should expect complete unacceptability for the reasons noted above. The content of a *Schema*-based prediction is that there is no such numeration. The informant judgment that $\alpha$ is not completely unacceptable under $\gamma(a, b)$ (even if not fully acceptable) would therefore disconfirm a *Schema*-based prediction because that would mean, contrary to the prediction, that there is numeration $\mu$ corresponding to $\alpha$ that would result in LF($\mu$) (hence SR($\mu$)) compatible with $\gamma(a, b)$ and PF($\mu$) (hence pf($\mu$)) non-distinct from $\alpha$—and $\alpha$ is not fully acceptable under $\gamma(a, b)$, it must be due to extra-grammatical

---

16 We will consider below two other possible formulations of an okSchema-based prediction.

Language-3-5-3.doc

7/50
factor(s). While the marginal acceptability would thus disconfirm a *Schema-based prediction, it would be compatible with, and hence would in fact confirm, an ok*Schema-based prediction as formulated in (7) or (8).  

Given that the ultimate testability of our hypotheses lies in their being subject to disconfirmation, what makes our hypotheses testable is the *Schema-based predictions they give rise to. To put it differently, it is most crucially by making *Schema-based predictions that we can seek to establish a "fact" that needs to be explained in research concerned with the properties of the Computational System and that serves as evidence for or against hypotheses about the Computational System. To ensure that the complete unacceptability of the *Examples is indeed due to the hypothesized grammatical reason, we must also try to demonstrate that (i) ok*Examples that minimal differ from the *Examples in terms of the hypothesized formal property are acceptable under γ(a, b) (at least to some extent) and (ii) ok*Examples that are identical to the *Examples but do not involve interpretation γ(a, b) are acceptable (at least to some extent). The acceptability of the former type of ok*Examples would indicate that the complete unacceptability of the *Examples under discussion cannot easily be attributed to the unnaturalness of the "meaning" of the entire sentence. The acceptability of the latter type of ok*Examples, on the other hand, would indicate that the complete unacceptability of the *Examples cannot be due to parsing difficulty.

Let us say that a predicted schematic asymmetry gets confirmed, i.e., a confirmed schematic asymmetry obtains, if and only if the informants' judgments on *Examples are consistently "completely unacceptable" and their judgments on the corresponding ok*Examples are not "completely unacceptable." By using the numerical values of "0" and "100" for "complete unacceptability" and "full acceptability," respectively, we can express what we intend as follows: a confirmed schematic asymmetry obtains if and only if the "representative value" of the *Schema is 0 and that of the corresponding ok*Schemata is higher than 0. The *Schema-based prediction in question must survive a rigorous disconfirmation attempt while at the same time the corresponding ok*Schema-based predictions must be confirmed. Otherwise, the predicted schematic asymmetry does not get confirmed. On the basis of the considerations given above, I would like to suggest that confirmed schematic asymmetries be regarded as "basic units of facts" for research concerned with the properties of the Computational System, i.e., as long as our research is concerned with the properties of the Computational System of the language faculty, our hypotheses should make predictions about, and be evaluated in terms of, confirmed schematic asymmetries.

As noted, while the requirement on the *Schema-based prediction is quite strict, how strict a requirement we

---

17 Among what is disregarded here in the somewhat simplified presentation in the text is the possibility that the informant does not fully understand the intended interpretation as provided in the instructions; the informant judgment in such cases could not be revealing about the properties of the Computational System under discussion. I will return to this and other related issues in section 4.2, where I address the significance of invoking γ(a, b).

18 As pointed out, fluctuation and variation are expected if we are considering an ok*Schema-based prediction but not if we are considering a *Schema-based prediction. When judgmental disagreement and fluctuation are observed, it must therefore be understood clearly whether we are considering a *Schema-based prediction or about an ok*Schema-based prediction. What is most crucial is that there be as little fluctuation and variation as possible in informant judgment with regard to *Schema-based predictions. The recognition of this point reduces the degree of murkiness of the relevant informant judgments considerably, and thereby resulting in a substantial increase of the testability of our hypotheses.

19 It is sometimes not possible to construct the latter type of ok*Examples. We can in such cases try constructing ok*Examples whose acceptability would indicate that the complete unacceptability of the corresponding *Examples is independent of parsing difficulty.

20 The "representative value" of a Schema is computed based on the informant judgments on the Examples that conform to the Schema; cf. Ueyama 2010 for more details. In what follows, the "average score" is sometimes used instead of the "representative value." In actual practice, we must allow some room for the possibility of "errors" committed by informants. We might therefore have to be "content" with something like "5 or less" or "around 5" as the "representative value" of the *Schema among the entire informants, on the scale of "0" (for complete unacceptability) to "100" (for full acceptability), for example. It must be understood that, if some informants consistently find *Examples of a given *Schema more or less acceptable, that should be regarded as a serious challenge to the hypotheses in question even if the "representative value" of the *Schema among the entire informants is quite low.
should impose on our \(\text{ok}\text{Schema}\)-based predictions may depend on various factors. It seems clear, however, that we cannot expect to convince others if the "representative value" of our \(\text{ok}\text{Schema}\) is 10, 20, or 30, for example, on the scale of "0" (for complete unacceptability) to "100" (for full acceptability), even if that of the corresponding *Schema is 0. While it is bound to be a subjective matter to determine what the "representative value" of the \(\text{ok}\text{Schemata}\) should be in order for a confirmed schematic asymmetry to obtain, the researchers themselves perhaps should aspire to the "standard" suggested in (9), leaving aside its actual feasibility in every single experiment.21

(9) An \(\text{ok}\text{Schema}\)-based prediction—version 3:
Informants judge every \(\text{ok}\)Example (in an experiment) conforming to an \(\text{ok}\text{Schema}\) to be fully acceptable under interpretation \(\gamma(a, b)\).

One might suggest that identifying confirmed schematic asymmetries is analogous to the rigorous observation and recording of the positions of planets made by Tycho Brahe and other such observations in physical sciences. Feynman (1965/94), for example, remarks as follows:22

... The ancients first observed the way the planets seemed to move in the sky and concluded that they all, along with the earth, went around the sun. This discovery was later made independently by Copernicus, after people had forgotten that it had already been made. Now the next question that came up for study was: exactly how do they go around the sun, that is, with exactly what kind of motion? Do they go with the sun as the centre of a circle, or do they go in some other kind of curve? How fast do they move? And so on. This discovery took longer to make. The times after Copernicus were times in which there were great debates about whether the planets in fact went around the sun along with the earth, or whether the earth was at the centre of the universe and so on. Then a man named Tycho Brahe evolved a way of answering the question. He thought that it might perhaps be a good idea to look very very carefully and to record exactly where the planets appear in the sky, and then the alternative theories might be distinguished from one another. This is the key of modern science and it was the beginning of the true understanding of Nature—this idea to look at the thing, to record the details, and to hope that in the information thus obtained might lie a clue to one or another theoretical interpretation. So Tycho, a rich man who owned an island near Copenhagen, outfitted his island with great brass circles and special observing positions, and recorded night after night the position of the planets. It is only through such hard work that we can find out anything.

When all these data were collected they came into the hands of Kepler, who then tried to analyse what kind of motion the planets made around the sun. And he did this by a method of trial and error. At one stage he thought he had it; he figured out that they went around the sun in circles with the sun off centre. Then Kepler noticed that one planet, I think it was Mars, was eight minutes of arc off, and he decided this was too big for Tycho Brahe to have made an error, and that this was not the right answer. So because of the precision of the experiments he was able to proceed to another trial and ultimately found out three things [i.e., Kepler's three laws of planetary motion, HH].” (Feynman 1965/94: 5-6)

21 It is worth point out that we can address across-informant repeatability meaningfully only if we have obtained within-informant repeatability (especially for the researchers themselves) where repeatability is understood in terms of a confirmed schematic asymmetry. The relevant issues that deserve further discussion include concepts such as informant resourcefulness, and single-informant and multiple-informant experiments.

22 One may point out that identifying confirmed schematic asymmetries is more theory-laden than Brahe's observation of the motion of the planets because the construction of a *Schema and that of the corresponding \(\text{ok}\)Schemata are based on various hypotheses, including those about properties of the Computational System, those about how a certain type of informant intuition arises based on certain properties at LF—which are here called bridging statements (see section 2.3)—, about what LF representation(s) can, cannot, or must correspond to a particular surface phonetic string, etc. Confirmed schematic asymmetries are thus perhaps closer to observations that have been replicated with the aid of various observation devices.
Given that "[i]t is only through such hard work that we can find out anything," it is clear that we should bring the utmost rigor to our attempt to identify what the "facts" are. Working with confirmed schematic asymmetries, I would like to suggest, is "the key of modern science" of the language faculty and it might as well be "the beginning of the true understanding" of the language faculty by means of the general scientific method in (1).

2.5. The significance of experimental results

Before turning to the discussion of empirical materials for illustration of the proposal, I would like to make one last point in relation to the significance of experimental results. Recall that a confirmed schematic asymmetry obtains if and only if the *Schema-based prediction has survived a rigorous disconfirmation attempt and at the same time the corresponding okSchema-based predictions have been confirmed.

Suppose that we have designed and conducted an experiment, the *Schema-based prediction has not been disconfirmed and, furthermore, the corresponding okSchema-based predictions have been confirmed. Does this mean that we are justified to conclude that we now have a confirmed schematic asymmetry? The fact that the result of a particular experiment is in harmony with the prediction(s) does not quite lead us to conclude that we have obtained a confirmed schematic asymmetry. As noted, what is predicted by a *Schema-based prediction is that informants judge any *Example (conforming to a *Schema) to be completely unacceptable under the specified interpretation. The researcher might have tried his or her best to construct the *Examples that are most natural and the easiest to parse for the intended interpretation but it is still possible that someone else can construct "better" *Examples of the *Schema that are acceptable (to some extent) under the specified interpretation.

Once the experimental results have obtained as predicted in his or her own experiment(s), the researcher should therefore invite other interested researchers to construct *Examples (as well as okExamples) and conduct their own experiments in accordance with the predicted schematic asymmetry. That is to say, having obtained the expected informant judgments in our own experiment(s) is merely a start in terms of our rigorous disconfirmation attempt. Other interested researchers are thus strongly encouraged to conduct experiments themselves on the basis of the predicted schematic asymmetry, making various adjustments on the lexical choices in the actual Examples, as allowed in the Schemata, doing the best they can to construct *Examples of the *Schema that are acceptable (to some extent) under the specified interpretation. Given the otherwise inherently murky status of evidence in language faculty science, such a rigorous disconfirmation attempt would make us hopeful that we are dealing with something that is a reflection of properties of the Computational System of the language faculty.

The prediction is that the *Examples of the *Schema are completely unacceptable under the specified interpretation no matter how much effort might be made to save them from complete unacceptability. If the average score of the *Schema-based prediction(s) is not close to zero in any such experiment, we should reconsider the validity of our hypotheses and the soundness of our experimental design; we would have to consider how such informant judgments arise. That should be our basic attitude if we are interested in discovering the properties of the Computational System of the language faculty in line with the general scientific method schematized in (1). If the *Schema-based prediction does not get disconfirmed in many such experiments, we will finally be in a position to conclude, with some confidence, that the *Schema-based prediction has survived a rigorous disconfirmation attempt, and to the extent that the corresponding okSchema-based predictions get confirmed, we can say, again with some confidence, that we have indeed obtained a confirmed schematic asymmetry, which, I maintain, is a 'minimal unit of facts' for research concerned with properties of the Computational System, as noted above.23, 24

23 It is ultimately up to an individual researcher how strict a standard s/he wishes to adopt for determining when a confirmed schematic asymmetry has obtained. The researcher's decision will affect how effectively s/he could learn from errors in subsequent research and experiments. If a researcher decides to go by a lenient standard, it will quickly become unclear what we could learn from an experiment on new predictions—regardless of whether the new *Schema-based predictions are disconfirmed—because the new predictions are made on the basis of hypotheses that might have been "accepted" despite the absence of any compelling empirical justification; see section 4.2.
As our research advances, we expect our confirmed schematic asymmetries to represent increasingly more general and abstract generalizations, and we will be seeking to deduce their explanations (i.e., hypotheses that account for them) from more basic and fundamental principles, approaching something that may deserve to be called a truly explanatory theory of the language faculty that is empirically grounded in confirmed schematic asymmetries. No matter how abstract our theory of the language faculty may become, its empirical consequences should remain expressible, ultimately, in terms of confirmed schematic asymmetries.

3. Illustration

3.1. Hypotheses about local anaphors in English

It has been observed at least since the mid-1960s that informant judgments are in accord with a general pattern as illustrated in (10).

(10) a. John recommended himself.
    b. *John thought that Mary had recommended himself.

Attempts have been made to express the contrast as a reflection of the Computational System, resulting in a hypothesis about the Computational System that has the effect in (11) and a hypothesis about the mental Lexicon of speakers of English as in (12), as discussed in Chomsky 1981.

(11) A [+A] category must have an antecedent in its local domain.

(12) Himself is marked [+A] in the mental Lexicon of speakers of English.

By defining "local domain" so as to ensure that in (13) NP2 is, but NP1 is not, in the local domain of NP3, the contrast in (10) is accounted for.25

(13) NP1 Verb [that NP2 Verb NP3]

That is to say, if one puts forth or accepts a hypothesis that expression $\alpha$ is marked [+A], one makes a testable prediction—as long as one also accepts something like (11) and the definition of "local domain" that has the effect noted above. One of the clearest predictions is that sentences containing [+A]-marked $\alpha$ are unacceptable if $\alpha$ is an embedded object and is interpreted as expressing the same individual(s) as the matrix subject. We can state the predicted schematic asymmetry as follows:

(14) a. $^\text{ok} \text{Schema}$
    NP V himself
    $\text{NP=himself}$

24 The research attitude advocated here is thus quite different from one that takes the presence of some contrast in the predicted direction between some examples for some speakers as constituting evidence in support of the hypotheses that give rise to the prediction under discussion (even when the *Examples are not judged consistently unacceptable by the informants). As argued above, the mere fact that such a contrast obtains does not in and of itself mean much at all for research concerned with the properties of the Computational System in line with the general scientific method schematized in (1); see section 2.3 for its conceptual basis. I might add in passing that if a *Schema does not specify anything about prosody or intonation, the prediction must be that *Examples conforming to the *Schema are completely unacceptable no matter what prosody/intonation might be used; see Miyagawa and Arikawa 2007: 652 (at the end of their section 3) for a remark that seems to be based on a rather different view.

25 The use of "NP" in place of "DP" is inconsequential in this paper.
As suggested above, what is predicted is a *schematic asymmetry*; more specifically, the prediction is that every example conforming to (14b) is judged to be completely unacceptable under the interpretation indicated in (14a) while examples conforming to (14a) and (14c) are judged (more or less) acceptable under the interpretations indicated in (14a) and (14c), respectively. We are not going to address in this paper how robust the informant judgments are on this predicted schematic asymmetry; we only note here that an informal survey conducted a few years ago suggests that they are fairly robust in accordance with (14).

### 3.2. Hypotheses about "local anaphors" in Japanese

#### 3.2.1. Hypotheses

In much of the generative research over the past 20 years, Japanese expressions such as *otagai*, *zibun-zisin*, and *kare-zisin* have been assumed (or claimed) to be marked [+A] in the sense noted in section 3.1, and they are called local anaphors in Japanese. The claim that *otagai*, *zibun-zisin*, and *kare-zisin* are local anaphors can be stated as in (15); see (12).²⁶

(15) Specifications in the mental Lexicon of speakers of Japanese:

a. Otagai is marked [+A].

b. Zibun-zisin is marked [+A].

c. Kare-zisin is marked [+A].

The properties of the Computational System are assumed to be universal, with the possible exception having to do with the so-called head parameter. The hypothesis about the Computational System having the effect in (11) is thus considered universal.

(11) A [+A] category must have an antecedent in its local domain.

A reasonable application to Japanese of the notion of "local domain" as understood in relation to (13) would lead us to accept that in (16) NP2 is, but NP1 is not, in the "local domain" of NP3.

(16) NP1-ga [NP2-ga NP3-{o/ni} Verb to] Verb

With the language-specific lexical hypotheses in (15) and the universal hypothesis in (11), along with the relevant articulation of "local domains" in Japanese, we make testable predictions.

#### 3.2.2. *Schema-based predictions and *okSchema*-based predictions

The predicted schematic asymmetries as indicated in (17) and (18) below are among the consequences of adopting (11), (15a), and the characterization of the "local domain" as noted above.

---

²⁶ I leave aside the issue as to whether and how each of (15) is derived from more basic statements; this applies to (12) as well. Many generative works dealing with Japanese have provided some paradigm or other in support of (15), and other works have derived and discussed various empirical as well as theoretical consequences by assuming the validity of the lexical hypotheses in (15). Such works are in fact numerous and they include Nishigauchi 1992, Saito 1992, 2003, and Miyagawa and Arikawa 2007.
Each Schema in (17) corresponds to something like "NP(1) thinks [that NP(1) Verb otagai/karera]" or "NP(1) wonders [why NP(1) Verb otagai/karera]," where NP1 (the embedded subject in (17a) and the matrix subject in (17b) and (17c)) is the intended "antecedent" of otagai or karera 'them' in the embedded object position. Because of (11), (15a) and the characterization of "local domains" in Japanese as discussed above, (17b) is a *Schema while (17a) is an okSchema. Because karera 'them' is not marked [+A], (17c) is also an okSchema.

In each Schema in (18) below, the intended antecedent for otagai or karera is the relative head (NP1), which is presumably related with the subject (ec) in the relative clause. As in the case of (17), (18b) is a *Schema while (18a) and (18c) are okSchemata.

On the basis of the Schemata in (17), we can construct the Examples in (19) and (20).27
c. *Example

John to Bill-wa [sensei-ga naze karera-o suisensi-ta no ka] mottaku wakara-nakat-ta
'John and Bill had no idea why the teacher had recommended them.'

On the basis of the Schemata in (18), we can construct the Examples in (21) and (22).

(21) a. *Example

'John and Bill, who had voted for each other at the election last week, were surprised to learn who Susan had voted for.'

b. *Example

'John and Bill, who thought that Susan had voted for each other at the election last week, were surprised to learn who Susan had voted for.'

c. *Example

'John and Bill, who thought that Susan had voted for them for the election last week, were surprised to learn who Susan had voted for.'

(22) a. *Example

'I hear that John and Bill, who had recommended each other for the new post, are emailing various people to explain why.'

b. *Example

'I hear that John and Bill, who wanted to know why Mike had recommended each other for the new post, are emailing various people to find out why.'

c. *Example

'I hear that John and Bill, who wanted to know why Mike had recommended them for the new post, are emailing various people to find out why.'

The predictions are thus as follows:

(23) The *Schema*-based prediction:

The *Examples* conforming to the *Schemata* in (17b) and (18b) are completely unacceptable, including the (b) examples in (19)-(22).

(24) The *Schema*-based prediction:

The *Examples* conforming to the *Schemata* in (17a), (18a), (17c) and (18c) are not completely unacceptable, including the (a) and (c) examples in (19)-(22).

Similarly, the schematic asymmetries as indicated in (25) and (26) are among the consequences of adopting (15b), along with (11) and the characterization of the "local domain" as noted above.
(25)  a. \(\text{\textit{okSchema}}\)
    \[\text{NP-ga/wa} \ [\text{NP1-ga zibun-zizin-o/ni V-ru/ta \{to/no ka\}} \] V-ru/ta \]
    \[\text{NP1} = \text{zibun-zisin}\]

    b. \(\text{\textit{*Schema}}\)
    \[\text{NP1-ga/wa} \ [\text{NP-ga zibun-zizin-o/ni V-ru/ta \{to/no ka\}}] \] V-ru/ta
    \[\text{NP1} = \text{zibun-zisin}\]

    c. \(\text{\textit{okSchema}}\)
    \[\text{NP1-ga/wa} \ [\text{NP-ga \{kare/kanozyo\}-o/ni V-ru/ta \{to/no ka\}} \] V-ru/ta \]
    \[\text{NP1} = \text{kare/kanozyo}\]

(26)  a. \(\text{\textit{okSchema}}\)
    \[\text{[[[\alpha-ga zibun-zisin-o/ni V-T to] V-T] NP]-wa} ... \]
    \[\alpha = \text{zibun-zisin}\]

    b. \(\text{\textit{*Schema}}\)
    \[\text{[[[NP-ga zibunzisin-o/ni V-T to] V-T] \alpha]-wa} ... \]
    \[\alpha = \text{zibun-zisin}\]

    c. \(\text{\textit{okSchema}}\)
    \[\text{[[[NP-ga \{kare/kanozyo\}-o/ni V-T to] V-T] \alpha]-wa} ... \]
    \[\alpha = \text{kare/kanozyo}\]

With (15c), we obtain the \textit{schematic asymmetries} as indicated in (27) and (28).

(27)  a. \(\text{\textit{okSchema}}\)
    \[\text{NP-ga/wa} \ [\text{NP1-ga kare-zizin-o/ni V-ru/ta \{to/no ka\}} \] V-ru/ta \]
    \[\text{NP1} = \text{kare-zisin}\]

    b. \(\text{\textit{*Schema}}\)
    \[\text{NP1-ga/wa} \ [\text{NP-ga kare-zizin-o/ni V-ru/ta \{to/no ka\}}] \] V-ru/ta
    \[\text{NP1} = \text{kare-zisin}\]

    c. \(\text{\textit{okSchema}}\)
    \[\text{NP1-ga/wa} \ [\text{NP-ga \{kare/kanozyo\}-o/ni V-ru/ta \{to/no ka\}} \] V-ru/ta \]
    \[\text{NP1} = \text{kare/kanozyo}\]

(28)  a. \(\text{\textit{okSchema}}\)
    \[\text{[[[\alpha-ga kare-zisin-o/ni V-T to] V-T] NP]-wa} ... \]
    \[\alpha = \text{kare-zisin}\]

    b. \(\text{\textit{*Schema}}\)
    \[\text{[[[NP-ga kare-zisin-o/ni V-T to] V-T] \alpha]-wa} ... \]
    \[\alpha = \text{kare-zisin}\]

    c. \(\text{\textit{okSchema}}\)
    \[\text{[[[NP-ga kare-o/ni V-T to] V-T] \alpha]-wa} ... \]
    \[\alpha = \text{kare}\]
The next subsection discusses only (15a) and its empirical consequences; see Appendix for some examples conforming to the schemata in (25)-(28) and the results of experiments on the predicted schematic asymmetries as indicated in (25)/(26) and (27)/(28).

3.3. Experiments

One can test *Schema-based predictions and corresponding *Schema-based predictions by checking informant judgments on *Examples and the corresponding *Examples, so as to see whether we obtain a confirmed schematic asymmetry. This subsection briefly introduces the general design of experiments that we have been conducting for the past few years along with the relevant experimental results.

The examples are presented on-line to the informants, along with the specification of their intended interpretation. The specifications of the intended interpretations are as in (29), for example, once translated into English.

(29) a. under the interpretation that "John voted for Bill and Bill voted for John"
   b. under the interpretation that karera 'them' and John to Bill 'John and Bill' refer to the same individuals

In an experiment on the predicted schematic asymmetries in (17) and (18), for example, the 12 Examples in (19)-(22) are presented to informants in a random fashion, (i) one at a time or (ii) three at a time (e.g., those in (19)), depending upon the test type chosen by each informant.

Depending upon the test type of their choice, the informants either (i) choose "No" (for "not acceptable no matter what") or "Yes" (for "(more or less) acceptable") or (ii) indicate how acceptable they find each example by clicking one of the five radio buttons as in (30). And what the informant has indicated is converted to numerical values as in (31), i.e., the worst score is converted to 0 and the best score to 100. Likewise, the "Yes" and the "No" answers in the "Yes-or-No" test get converted to 0 and 100, respectively.

(30) Bad < ===== > Good
     o o o o o o

(31) 0, 25, 50, 75, 100

According to the results we have obtained so far, the choice of the "test type" does not make a significant difference.

The informants are allowed to return to the experiment website and report their judgments on examples in the same experiment as many times as they wish; they may repeat the same "test type" as before or try a different "test type" (as to "Yes-or-No" or "Five-ranking" and also as to "one at a time," or "three at a time"). In the event that one informant has reported his/her judgment on the same experiment more than once (regardless of the "test type"), the average of that informant's judgments on a given example is used in calculating the mean score on the example by the entire informants.

(32) is a summary of the results of the experiment on the predicted schematic asymmetries in (17) and (18).

---

28 The informants are also given the option of not providing their judgments on an example.
A summary of the results of the experiment as of June 14, 2010.\(^{29}\)

<table>
<thead>
<tr>
<th>Schema group 1</th>
<th>Otagai is in the embedded object position.</th>
<th>Schema 1 A (=((17a)))</th>
<th>68 values</th>
<th>96*Examples in (19a) and (20a)</th>
<th>Schema 1 B (=((17b)))</th>
<th>67 values</th>
<th>61*Examples in (19b) and (20b)</th>
<th>Schema 1 C (=((17c)))</th>
<th>68 values</th>
<th>85*Examples in (19c) and (20c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schema group 2</td>
<td>Otagai is in the embedded object position. The intended antecedent is the relative head.</td>
<td>Schema 2 A (=((18a)))</td>
<td>68 values</td>
<td>94*Examples in (21a) and (22a)</td>
<td>Schema 2 B (=((18b)))</td>
<td>68 values</td>
<td>53*Examples in (21b) and (22b)</td>
<td>Schema 2 C (=((18c)))</td>
<td>68 values</td>
<td>70*Examples in (21c) and (22c)</td>
</tr>
</tbody>
</table>

34 informants, 865 answers

<table>
<thead>
<tr>
<th>Schema Group 1</th>
<th>Schema A</th>
<th>Schema B</th>
<th>Schema C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schema Group 1</td>
<td>68 values</td>
<td>96</td>
<td>67 values</td>
</tr>
<tr>
<td>Schema Group 2</td>
<td>68 values</td>
<td>94</td>
<td>68 values</td>
</tr>
<tr>
<td>Total</td>
<td>136 values</td>
<td>95</td>
<td>135 values</td>
</tr>
</tbody>
</table>

"Schema group 1" is for (17) and "Schema group 2" is for (18). "Schema 1 A" covers the *Examples in (19a) and (20a), "Schema 1 B" the *Examples in (19b) and (20b), and "Schema 1 C" the *Examples in (19c) and (20c). Likewise, "Schema 2 A" covers the *Examples in (21a) and (22a), "Schema 1 B" the *Examples in (21b) and (22b), and "Schema 1 C" the *Examples in (21c) and (22c). Of the 34 informants, 14 are "linguistically naïve" and 20 are "linguistically informed."\(^{30}\) "865 answers" means that there have been 865 occurrences of a reported judgment. Some informants have judged the same example more than once; but in such cases the values in the summary chart in (32) are based on the average score on a given example by each such informant. As noted, corresponding to each Schema in the experiment in question, there are two Examples. The *Examples in (19b) and (20b), for instance, correspond to the *Schema in (17b) (i.e., Schema 1B in (32)). The average scores on (19b) and (20b) are 50 and 72, respectively, and "61" for Schema 1B in (32) is the average of those two scores.

The average scores of "Schema 1 B" and "Schema 2 B" should be close to 0 according to the *Schema-based predictions in line with the predicted schematic asymmetries in (17) and (18). The informant judgments as summarized in (32) thus clearly disconfirm the *Schema-based predictions based on the lexical hypothesis in (15a), repeated here.\(^{31, 32}\)

---

\(^{29}\) See section 1 of the supplementary file for more details of the experimental results.

\(^{30}\) When registering for the on-line experiments, informants are asked several questions, including one about their dialects. They are also asked whether they understand (i) "bound variable anaphora" or "bound readings" and (ii) "A takes wide scope over B" as they are used in linguistic discussion. If they state that they understand at least (i) or (ii), they are "classified" as "linguistically informed" for the purpose of the discussion in this paper. If they state that they understand neither, on the other hand, they are "classified" as "linguistically naïve." The classification in question is thus based on what each informant "declares" him/herself. Informant resourcefulness, at least with respect to a particular (set) of experiment(s), can be measured on the basis of the judgments that the informant report in preliminary experiments.

\(^{31}\) "Linguistically-naïve" informants (14 informants) tend to judge the *Examples somewhat less acceptable than "linguistically-informed" informants (20 informants) (e.g., 39 by the former and 70 by the latter on Schemata B in (32)); but the former also judge *Examples less acceptable than the latter (e.g., 63 by the former and 88 by the latter on Schemata C in (32)). This is expected from the considerations of informants' resourcefulness.
(15) a. *Otagai* is marked [+A].

It may be possible that someone can in the future come up with a way to modify (15a) and hence save a version of (15a). It is important to note that if such an attempt results only in the elimination of the *Schema*-based prediction without introducing a new *Schema*-based prediction or if it results in the disconfirmation of the new *Schema*-based prediction, that will be content-reducing (or degenerating) problemshift in the terms of Lakatos 1970/1978. I leave the challenge of saving (15a) (and for that matter (15b) and (15c) as well) in a theoretically or empirically progressive way to those who wish to make use of those language-specific hypotheses in their theoretical discussion.33

While it is not possible to empirically demonstrate the non-existence of elements in Japanese that are marked [+A]—for it is not possible to empirically demonstrate the non-existence of anything—their non-existence in Japanese is an immediate consequence if we adopt the thesis put forth in Fukui 1986. Fukui (1986) proposes that the mental Lexicon of speakers of Japanese does not contain what is responsible for making functional categories "active." Given the assumption that what most crucially underlies a local anaphor is an "active functional category"—cf. Lebeaux 1983 and Chomsky 1986: 175f—the absence of local anaphors in Japanese is as expected,

32 There is no space for a detailed discussion, but Japanese *Examples such as those corresponding to "each other's lovers tried to seduce *John and Bill*" and "the warm spring breeze made each other feel very happy" have been judged acceptable in other experiments; see (48). The mean "scores" on these two types of *Examples, as of June 14, 2010, are 60 (by 20 informants) on the former and 93 (by 19 informants) on the latter, contrary to the predicted 0. Likewise, the *Schema*-based predictions in accordance with the two other lexical hypotheses in (15b) and (15c) have also been disconfirmed, very much like the way the *Schema*-based predictions under the lexical hypothesis in (15a) have been disconfirmed; see Appendix.

As H.-D. Ahn (p.c., 12/12/2009) suggests, one might pursue the possibility that the hypotheses in (15) are valid but that *otogai, zibun-zisin and kare-zisin* always occur in a structural position in which they have a covert antecedent in its local domain. Such a move would save (15) from refutation (and one might even claim that it would also make it possible to maintain the thesis that Japanese shares a "universal property" of having [+A] elements in its Lexicon), but it would result in the elimination of the *Schema*-based predictions. Hence that would be like a content-reducing (or degenerating) problemshift in the terms of Lakatos 1970/1978 unless it were accompanied by the introduction of a new *Schema*-based prediction. Furthermore, if we accepted the view that what formally underlies a local anaphor is something like an "active functional category," it would be puzzling that there does not seem to be any confirmed schematic asymmetry in support of the presence of an "active functional category" in Japanese; see the remarks in the last paragraph of this section. I am not aware of any empirical evidence in support of the existence of DPs in Japanese and of the EPP (feature) in Japanese that forms a confirmed schematic asymmetry in accordance with the criteria advocated here.

33 One might point out that the "values" of B Schemata are significantly lower than those of the A and C Schemata in (32) and suggest that the contrast can be taken as evidence in support of (15a). Suppose that the contrast in question obtains in experiment after experiment. It is reasonable to conclude that the contrast is not by chance and is a reflection of something. But it is not obvious that it is a reflection of the interaction between the universal property hypothesized in (11) and the Japanese-particular property hypothesized in (15a). Recall that the *Schema*-based prediction under discussion is that *Examples conforming to the *Schemata in (17b) and (18b) are completely unacceptable; see section 2.3. One could thus be justified to consider the contrast between the B Schemata and the A/C Schemata in (32) as evidence in support of (15a) only if one (i) identified the factors that are responsible for the informants' not judging the *Examples in (19)-(22) completely unacceptable and (ii) demonstrated that, by avoiding the effects of such factors, we could obtain a confirmed schematic asymmetry in a modified experiment.

In this connection, it may be worth noting the following: Of the 34 informants in the experiment whose results are summarized in (32), only two have judged the *Examples (almost) completely unacceptable. One of them is "linguistically informed" and the other is "linguistically naïve"; see footnote 30. Their values on the B Schemata in (32) are 0 and 3. (The other 32 informants accept *Examples of the *Schemata in (32) to varying degrees.) For one of them (the "linguistically-informed" one), the values on Japanese examples corresponding to "each other's lovers tried to seduce *John and Bill*" and "the warm spring breeze made each other feel very happy" are 87 and 75, respectively. See section 7 of the supplementary file. Hence, one cannot maintain that (15a) is valid for that informant. The other informant has not participated in the experiments that contain such examples.
and hence, the results of the Experiments reported and alluded to above are also as expected; cf. Narita 2010 for some relevant discussion. That is to say, the fact that the researchers have so far failed to identify what qualifies as a local anaphor in Japanese despite the concerted efforts by a substantial number of practitioners for nearly three decades, is not puzzling, after all.

4. Two research heuristics

4.1. Maximize testability

Nakaya (1958: 17) remarks that science has its intrinsic limitation such that it is a discipline where we extract reproducible phenomena and analyze them statistically. How do we go about extracting reproducible phenomena with regard to the language faculty? I have suggested that we seek reproducible phenomena by trying to establish confirmed schematic asymmetries, and proposed that confirmed schematic asymmetries be regarded as "basic units of facts" for research that aims at discovering the properties of the language faculty by the general scientific method in (1). In order to attain testability of our hypotheses, it is necessary to have a means to identify "facts" to be accounted for by our hypotheses. Without such a means, we could not tell the exact empirical content of our predictions (i.e., what we deduce from our hypotheses) and hence could not compare them with the results of our experiments; see section 1. One way to maximize testability in our research is therefore to try to identify, and work on, 'phenomena' that (are likely to) lead us to confirmed schematic asymmetries.

There is an additional means to maximize testability. A confirmed schematic asymmetry obtains if and only if (i) a *Schema-based prediction has survived a rigorous disconfirmation attempt and (ii) the corresponding Schema-based predictions have been confirmed. As discussed in section 2.4, it is most crucially by making *Schema-based predictions that we can seek to attain testability of our hypotheses. We can thus try to maximize testability by pursuing hypotheses that give rise to as many *Schema-based predictions as possible. That is to say, the maximization of testability can be pursued not only by choosing to investigate certain phenomena but also by pursuing their accounts such that they lead to as many *Schema-based predictions as possible.

Let us record these two aspects of the "Maximize testability" heuristic.

(33) The two aspects of the "Maximize testability" heuristic:

34 I will return to the use of statistics in section 6.1.

35 The only structure-building operation assumed in the Computational System is (internal and external) Merge. In an attempt to identify informant intuitions that are a likely reflection of properties of the Computational System, we can thus reasonably focus on the informant intuitions that arise only on the basis of a structural relation created by Merge, i.e., only on the basis of a c-command relation. Given that the semantic interpretations that are based on structural properties of the Computational System are based on an LF representation, in accordance with the model of judgment making adopted here, it follows that we should focus on informant intuitions that arise only on the basis of a c-command relation at LF. That is in fact what is proposed in Reinhart 1983, in slightly different terms. LF representations are, of course, not observable. And that is why we must commit ourselves to hypotheses about correspondences between a pf string in a given language and an LF representation, without which we would not be able to make any testable predictions in language faculty science.

36 See section 2.4 for the importance of *Schema-based predictions.

37 One may observe that it is not clear exactly how we can measure the degree of maximization in question because one can increase the number of *Schema-based predictions in a non-trivial manner, by counting various sub-cases of a more general *Schema-based prediction as "distinct" *Schema-based predictions, for example. In the present work, I do not discuss this issue further, only noting that, when we study actual work, it is not difficult to see whether efforts are made to yield *Schema-based predictions or whether the work is mainly concerned with avoiding disconfirmation of a *Schema-based prediction (if it addresses *Schema-based prediction at all).
a. We should work on 'phenomena' that (are likely to) lead us to confirmed schematic asymmetries.
b. We should pursue hypotheses that give rise to as many *Schema-based predictions as possible, in a
non-trivial manner; see footnote 37.

I suggest that, when one pursues a particular hypothesis as part of an account of a given confirmed schematic asymmetry, one must try to adhere to (33b), not only with regard to the language being directly dealt with but also with regard to other languages (and in fact any other language); in other words, one must ask oneself whether and how a confirmed schematic asymmetry in a particular language and one's account of it could lead to new *Schema-based predictions that are in principle testable in any language and proceed in accordance with (33b).38

I should like to note that (33b) is also applicable when we modify our hypotheses in response to the failure of our predictions; i.e., modification of our hypotheses should proceed in line with (33b). If the modification only resulted in the elimination of a *Schema-based prediction without introducing a new *Schema-based prediction, that would be like a content-reducing (i.e., degenerating) problemshift in the terms of Lakatos 1970/1978; see the end of section 3.3.

4.2. Maximize our chances of learning from errors39

In "re-stating all the controversial things [he has] been saying in a number of theses," Popper 1963: 965 states that "[t]he growth of knowledge, and especially of scientific knowledge, consists in learning from our mistakes" and that "[w]hat may be called the method of science consists in learning from our mistakes systematically; first, by daring to make mistakes ...; and second, by searching systematically for the mistakes we have made, that is, by the critical discussion and the critical examination of our theories."40

How can we learn from our errors in research concerned with the language faculty? It is explicitly stated in Duhem 1909/1954: chapter 6, section 2: 185, and it has subsequently been widely agreed upon, that "if the predicted phenomenon is not produced, ... [t]he only thing the experiment teaches us is that among the propositions used to predict the phenomenon and to establish whether it would be produced, there is at least one error; but where this error lies is just what it does not tell us."41 We can try to maximize our chances of learning

38 It is perhaps worth reminding ourselves that, for research that aims at discovering the properties of the Computational System, language-specific hypotheses, such as (15a) would be significant only insofar as they contributed to a discovery of properties of the Computational System through their interaction with universal hypotheses (about the Computational System) such as (11).

39 A more appropriate phrasing of what is intended here may be something like "Making learning from errors possible and maximize the empirical nature of the mode of learning from errors." Among the reasons for wanting to avoid "chances" is that "chance" has, for many people, a probability-related meaning, as pointed out to me by Yasuo Deguchi (p.c., March, August 2010). In this paper, I only make this qualification here and continue to use what is given as the heading of this subsection. There are various related issues that are worth discussing, but they will have to be addressed in separate works.

40 These are two of the 17 points (the second and the third) Popper mentions. The first, fourth, and fifth points are also of direct relevance to the present discussion and they are reproduced here.

(i) (Popper 1963: 965, (1), (4), and (5))
   a. All scientific knowledge is hypothetical or conjectural.
   b. Among the most important arguments which are used in this critical discussion are arguments from experimental tests.
   c. Experiments are constantly guided by theory, by theoretical hunches of which the experimenter is often not conscious, by hypotheses concerning possible sources of experimental errors, by hopes or conjectures about what will be a fruitful experiment—which means by theoretical hunches that experiments of a certain kind will be theoretically fruitful.

41 This is a consequence of the thesis that no testable consequences are deducible from a single hypothesis. Contrary to what is commonly understood in relation to the so-called Duhem-Quine thesis, Duhem restricts his thesis to physics; see Ariew (1984) for how "Duhem's thesis is not the Duhem-Quine thesis," which is the title of its section 1.
from errors by minimizing the number of hypotheses whose validity is to be directly tested in a given experiment, and more importantly, by not using hypotheses that have been shown to be invalid in earlier experiments.

The hypothesis in (15a) has been shown to be invalid in an experiment that tests the predictions made under (15a), (11) and the hypotheses that ensure (34).42

15) a. *Otagai* is marked [+A].

11) A [+A] category must have an antecedent in its local domain.

34) NP1 is not in the local domain of NP3 in the representation corresponding to (16).

16) NP1-ga [NP2-ga NP3-{o/ni} Verb to] Verb ‘NP1 Verb that NP2 Verb NP3’

In accordance with Duhem's thesis noted above, the disconfirmation of the *Schema*-based prediction thus cannot be attributed solely to (15a).

Notice, however, that the validity of what is intended by (11) is widely accepted, presumably on the basis of what would amount to *confirmed schematic asymmetries* in a number of languages, although there may be disagreements concerning exactly how to express (11) in theoretical terms (and how robust the reported judgments are)33. Similarly, the validity of (34) is widely accepted among the researchers who address how (11) applies to Japanese. The validity of both (11) and (34) thus allows us to reasonably attribute the disconfirmation of the *Schema*-based prediction to the invalidity of (15a).

If we had obtained a *confirmed schematic asymmetry* in accordance with the predictions made under (15a), combined with (11) and (34), we would have acquired a new tool, so to speak, for our further probe into properties of the Computational System (and also into Japanese-particular properties, for that matter), because we could in that case assume (15a) to be valid in our further investigation and experiments. But we did not; i.e., we did not obtain a *confirmed schematic asymmetry* in accordance with the predictions made under (15a), combined with (11) and (34).

Now that the *Schema*-based predictions made under (15a) have been disconfirmed, it follows from the above reasoning that it is perhaps premature, and even ill-advised, to examine further predictions we might make on the basis of (15a) and other (new) hypotheses. Suppose that one combined (15a) with other new hypotheses (language-particular and/or universal) and made a new *Schema*-based prediction. It would not be clear what significance we could assign to the result of an experiment on the new "predicted" schematic asymmetry. If the new *Schema*-based prediction got disconfirmed, we could not attribute it to the newly-introduced hypothesis/ses because we independently know that (15a) is not valid, to begin with. Even if we obtained a *confirmed schematic asymmetry* we would not be justified to consider the newly-introduced hypothesis/ses and the "original hypothesis" in (15a) to be both valid because we already know that the latter is not. The results of the new experiment thus would not tell us anything about the newly-introduced hypothesis/ses, regardless of whether the new *Schema*-based prediction gets disconfirmed, because of the use of the hypothesis in (15a), which has already been invalidated.44

42 There are actually other hypotheses involved, such as those having to do with the model of the Computational System, including its existence, and the structure-building operation postulated in the Computational System (internal and external Merge); but they are part of the hard core of the research program in the terms of Lakatos 1970/1978, not being subject to (direct) empirical invalidation.

43 See the end of section 2.1.

44 See Poincaré 1952: chap. 9, 151-152 for relevant discussion. The general point made in the text seems to be rather poorly understood in the field (at least in works dealing with Japanese syntax), judging from the continued use of various types of hypotheses that have been shown not to be backed up by a *confirmed schematic asymmetry*, not only in regard to "local
In the preceding discussion, we have focused on the Subject-Object-Verb order (SOV). Given that the Object-Subject-Verb order (OSV) is more "complex" than SOV, as most practitioners in generative grammar dealing with Japanese seem to assume, it thus follows that there would be little merit to considering predictions about OSV made under (15a). Such predictions would be made under some hypotheses about the properties of the OSV as well as the hypothesis in (15a); but (15a) has already been shown to be invalid. The use of (15a) in addressing issues pertaining to more "complex" cases of OSV—including so-called long-distance OSV or multiple OS constructions (i.e., so-called multiple scrambling constructions)—would be even more problematic. A general research heuristic is therefore that if a testable hypothesis is shown to be invalid in what is considered to be a "simple case," we must avoid using it for what we consider to be a "more complex case." This is in line with the "Maximize our chances of learning from errors" heuristic.45

The considerations just given are applicable not only to research dealing directly with Japanese but also to cross-linguistic research that makes reference to Japanese-particular hypotheses such as (15a). Suppose that some cross-linguistic research is crucially based on the validity of the hypothesis in (15a) in Japanese or some other hypothesis that has been shown—by the disconfirmation of the *Schema-based prediction in question—not to be backed up by a confirmed schematic asymmetry. If a new prediction is made about a language other than Japanese crucially on the basis of such a hypothesis, it is not clear what significance can be assigned to the result of a new experiment, regardless of its outcome, for the reason noted above.

The most general research guideline of the methodology suggested here is the "Maximize our chances of learning from errors" heuristic. Ensuring and maximizing testability is a necessary condition for abiding by this heuristic. For, without ensuring testability, our predictions could not be disconfirmed and hence we would not have the chance to learn from errors. We can attain testability of our hypotheses most effectively by making reference to confirmed schematic asymmetries, and most crucially to *Schema-based predictions, provided that the ultimate testability of our hypotheses lies in their being subject to disconfirmation; see section 2.4.46 Our emphasis on the importance of building our research on confirmed schematic asymmetries and most crucially on *Schema-based predictions thus derives from the "Maximize our chances of learning from errors" heuristic.

In the foregoing discussion, I have addressed confirmed schematic asymmetries involving the interpretation γ(a, b), where an expression a is dependent upon another expression b for its interpretation; see footnote 9. I have noted in footnote 11 that it would be qualitatively more difficult to maximize our chances of learning from errors if we dealt with simple (un)acceptability without involving γ(a, b). And that means that we should first try to obtain confirmed schematic asymmetries involving γ(a, b) before we try to obtain confirmed schematic asymmetries without involving γ(a, b) if we are to maximize our chances of learning from errors. In other words, we should first work on "phenomena" that involve some dependency interpretation before we work on "phenomena" that do not involve such interpretation. This may appear paradoxical in light of the common perception that the simple (un)acceptability of a given sentence is more manageable and simpler to deal with than the (un)availability of a particular interpretation in a given sentence. Before we proceed to further illustrate the proposed methodology in the next section, I would therefore like to make brief remarks on why the involvement

anaphors" but also in regard to variable binding, quantifier scope, "floating quantifiers," and other "empirical domains." The illustration of this, however, will have to be made on a separate occasion due to space considerations. For some relevant discussion, see Hoji 2003, 2006a and Mukai to appear.

45 One might suggest that abiding by this research heuristic may (severely) lower "research productivity" and it may not be realistic to do so, especially for those who still do not have a secure job in the field. This is actually a fairly common reaction to the research methodology suggested here. There are a number of issues one can address in relation to this. Although such issues are mostly sociological, there are nonetheless issues that deserve serious attention, as might have to do with the goal of one's research program, the scientific nature of one's research program, including whether one would (aspire to) consider one's research program as an exact science, in which one deduce point-value predictions and expect them to be supported experimentally. Discussing the relevant issues would make the paper too long and also would take us too afield from the main point of this paper. I am thus not going to discuss the issues any further here; I hope to address them in separate works.

46 Recall that *Schema-based predictions can, but Schema-based predictions cannot, be disconfirmed.
of $\gamma(a, b)$ is crucial for the purpose of maximizing our chances of learning from errors.

Let us consider what we can conclude from the informant judgment in each of the four cases in (35), where, as before, $\alpha =$ the presented sentence and $\gamma(a, b) =$ a dependency interpretation holding between $a$ and $b$.

$$\begin{array}{|c|c|}
\hline
\text{Informant Judgment} & \text{On the acceptability of } \alpha \\
& \text{under } \gamma(a, b) \\
\hline
\text{a. Acceptable (to some extent)} & \text{c. Acceptable (to some extent)} \\
\text{b. Completely unacceptable} & \text{d. Completely unacceptable} \\
\hline
\end{array}$$

The judgment in (35a), one might think, indicates that the informant has been able to come up with a numeration that would result in (i) a PF representation non-distinct from $\alpha$ and (ii) an LF representation in which the hypothesized necessary condition for $\gamma(a, b)$ is satisfied. But such thinking is warranted only if the informant understands what is meant by $\gamma(a, b)$ and judges $\alpha$ under $\gamma(a, b)$ as intended by the researcher. $^{47}$ We can try to maximize the likelihood that the informant indeed understands $\gamma(a, b)$ by conducting a preliminary experiment, such as those involving a lexical, but not a structural, condition for $\gamma(a, b)$, as will be discussed in section 5.2. The reported judgments in such experiments can reveal whether the informants indeed understand what is meant by $\gamma(a, b)$, in addition to how well they understand the "instructions." If the same informant consistently rejects Examples of a certain schema (i.e., *Examples) that involves the same $\gamma(a, b)$, that also makes it reasonable for us to regard the informant judgment in (35a) as indicating that the informant understand what is meant by $\gamma(a, b)$ and has been able to come up with a numeration that results in (i) a PF representation non-distinct from $\alpha$ and (ii) an LF representation in which the hypothesized necessary condition for $\gamma(a, b)$ is satisfied. We thus have a reasonably reliable basis for regarding the judgment in (35a) as disconfirmation of the relevant *Schema-based prediction, and most importantly, as evidence against the hypotheses that have given rise to the prediction.

Turning now to the informant judgment in (35c), which is on simple (un)acceptability, it is not clear how we can make sure that such judgment arises because the informant has been able to come up with a numeration that would result in a PF representation non-distinct from $\alpha$. For it is not clear how we can make sure that the judgment in (35c) reflects "grammaticality" not "intelligibility." The judgment in (35c) might be due to the informant's finding $\alpha$ intelligible. As noted, in the case of (35a), we can turn to a preliminary experiment involving a lexical, but not structural, condition for $\gamma(a, b)$, in an attempt to make sure that the judgment in (35a) is not based on mere intelligibility, and we can also check the same informant's judgments on other examples involving the same $\gamma(a, b)$. But that is not an option in the case of (35c) because we are not considering $\gamma(a, b)$. The informant judgment in (35c) thus cannot be taken as evidence against the hypotheses that have given rise to the *Schema-based prediction in question, unlike the informant judgment in (35a). We are thus led to conclude that we do not have rigorous testability (refutability) of our hypotheses if we deal only with simple (un)acceptability.

One might take the judgment in (35b) as indicating that the informant has not been able to come up with a numeration that would result in (i) a PF representation non-distinct from $\alpha$ and (ii) an LF representation in which the hypothesized necessary condition for $\gamma(a, b)$ is satisfied. It is, however, possible that there is such a numeration but the informant just has not been able to come up with it for various reasons. For example, if $\alpha$ may be too difficult for the informant to parse, at least within the limit of her/his patience or willingness to check various possibilities for "parsing" $\alpha$. It is also possible that the informant reports that $\alpha$ is completely unacceptable under $\gamma(a, b)$ because the informant finds the interpretation of the entire $\alpha$ extremely unnatural, quite independently of whether $\gamma(a, b)$ is in principle possible in. The reported unacceptability in such cases therefore cannot be taken as evidence for the failure of the condition for $\gamma(a, b)$ to be satisfied in any LF representation.

$^{47}$ It is also necessary that the informant pays close enough attention to $\alpha$. This is one of the necessary conditions for the informant judgments to be reliable in our experiments in general. See footnote 17.
corresponding \( \alpha \). It is for this reason that we have two \( ^{ak}\text{Schema} \) corresponding to a \( ^{ak}\text{Schema}, \) as discussed in section 2.4. One \( ^{ak}\text{Schema} \) is meant to ensure that the informant accept \( ^{ak}\text{Example} \) which contain \( \gamma(\alpha, b) \) and whose entire interpretations are quite similar, if not identical, to that of the \( ^{ak}\text{Example} \) in question. The other \( ^{ak}\text{Example} \) is meant to ensure, as much as possible, that the reported unacceptability is not due to parsing difficulty. \( ^{ak}\text{Example} \) conforming to the first type of \( ^{ak}\text{Schema} \) thus contain the same \( \gamma(\alpha, b) \) that is crucially included in the interpretation of \( \alpha \), and those conforming to the latter is (as) identical to \( \alpha \) (as possible), but not involving \( \gamma(\alpha, b) \).

By obtaining the informant judgment that such \( ^{ak}\text{Example} \) are acceptable (at least to some extent), we will have a reasonable basis for assuming that the informant judgment that the corresponding \( ^{ak}\text{Example} \) is completely unacceptable is not due to parsing difficulty or the unnaturalness of the interpretation of the entire \( ^{ak}\text{Example}. \) And that makes us hopeful that the unacceptability judgment in question is indeed due to there being no numeration that would result in (i) a PF representation non-distinct from \( \alpha \) and (ii) an LF representation in which the hypothesized necessary condition for \( \gamma(\alpha, b) \) is satisfied. Recall that what is being tested is the validity of the empirical consequence, deduced from our hypotheses, that sentence \( \alpha \) is completely unacceptable under \( \gamma(\alpha, b) \) because the necessary condition for \( \gamma(\alpha, b) \) would not be satisfied in any LF representation corresponding to \( \alpha \). It is by invoking \( \gamma(\alpha, b) \) that we can reasonably consider the informant judgment in (35b) as arising because there is no numeration that would result in (i) a PF representation non-distinct from \( \alpha \) and (ii) an LF representation in which the hypothesized necessary condition for \( \gamma(\alpha, b) \) is satisfied.

Consider now the informant judgment in (35d), i.e., that \( \alpha \) is completely unacceptable, without involving \( \gamma(\alpha, b) \). As in the case of (35c), the unacceptability judgment in (35d) might be independent of what is hypothesized to be responsible for \( \alpha \)'s unacceptability. The reported unacceptability might be due to parsing difficulty or due to the unnaturalness of the interpretation of the entire \( \alpha \). We cannot, for example, exclude the possibility that parsing difficulty is responsible for the complete unacceptability of \( ^{ak}\text{Example} \) under a \( ^{ak}\text{Schema}-\text{based prediction} \) without involving \( \gamma(\alpha, b) \) because, unlike under a \( ^{ak}\text{Schema}-\text{based prediction involving} \gamma(\alpha, b) \), we cannot construct any \( ^{ak}\text{Example} \) that are phonetically identical to the \( ^{ak}\text{Example}; \) see section 2.4. In the case of (35c), we have a means to (try to) eliminate such possibilities because of the involvement of \( \gamma(\alpha, b) \). In the case of (35d), on the other hand, we do not have such a means. The informant judgment in (35d) therefore cannot be regarded as evidence that there is no numeration that would result in (i) a PF representation non-distinct from \( \alpha \) and (ii) an LF representation, and hence, as evidence for the hypotheses from which the relevant \( ^{ak}\text{Schema}-\text{based prediction} \) has been deduced. In short, for the purpose of discovering properties of the Computational System, it is not clear what significance we can assign to a \( ^{ak}\text{Schema}-\text{based prediction} \)'s surviving a (rigorous) disconfirmation attempt when \( \gamma(\alpha, b) \) is not involved, unlike when \( \gamma(\alpha, b) \) is involved.

The disconfirmation of a \( ^{ak}\text{Schema}-\text{based prediction involving} \gamma(\alpha, b) \) would tell us where to look. This is because a \( ^{ak}\text{Schema}-\text{based prediction involving} \gamma(\alpha, b) \) is made most crucially based on a hypothesized structural condition for \( \gamma(\alpha, b) \), coupled with independent hypotheses about the correspondences between phonetic strings in a particular language and LF representations—let us call such correspondences pf-LF correspondences.\(^{48}\) The content of a \( ^{ak}\text{Schema}-\text{based prediction} \) is that, corresponding to any \( ^{ak}\text{Example} \) conforming to a \( ^{ak}\text{Schema}, \) there is no numeration that would yield an LF representation in which the structural condition for \( \gamma(\alpha, b) \) would be satisfied. Recall that \( \alpha \) itself should be acceptable (more precisely, not completely unacceptable) if we did not consider \( \gamma(\alpha, b) \) (but see footnote 19). Hence, there should be a numeration that would yield a PF representation non-distinct from \( \alpha \) although no such numeration would result in an LF representation in which the condition for \( \gamma(\alpha, b) \) would be satisfied. The disconfirmation of a \( ^{ak}\text{Schema}-\text{based prediction involving} \gamma(\alpha, b) \), as in the case of the informant judgment in (35a) thus means that something is wrong with our hypotheses about condition for \( \gamma(\alpha, b) \) and/or the pf-LF correspondences we adopt (leaving aside the possibility that something is wrong with independent aspect(s) of the experimental design). The disconfirmation of a \( ^{ak}\text{Schema}-\text{based prediction involving} \gamma(\alpha, b) \) thus contributes to the maximization of our chances of learning from errors, contrasting sharply with the disconfirmation of a \( ^{ak}\text{Schema}-\text{based prediction} \) without involving \( \gamma(\alpha, b) \), as discussed above.

\(^{48}\) We are suppressing lexical hypotheses here.
One might wonder, in light of the preceding considerations, whether language faculty science cannot deal with "phenomena" that do not involve $\gamma(a, b)$. I would like to suggest that experiments on the simple (un)acceptability of a sentence $\alpha$ can be effectively carried out, i.e., confirmed schematic asymmetries without involving the "second at Schema" can be considered as "basic units of facts"—in line with the "Maximize our chances of learning from errors" heuristic—if the relevant research is based, directly or indirectly, on confirmed schematic asymmetries involving the interpretation $\gamma(a, b)$, where an expression $a$ is dependent upon another expression $b$ for its interpretation. That is to say, research that deals with $\gamma(a, b)$ (more specifically dependency interpretation $\gamma(a, b)$ that is crucially based on LF c-command (see footnote 35)) is more basic than research that deals with "phenomena" not involving such $\gamma(a, b)$.49

5. Further illustration: Confirmed schematic asymmetries in Japanese

5.1. A confirmed schematic asymmetry involving bound variable anaphora

The proposed methodology has been illustrated so far in relation to the language-particular lexical hypothesis that otagai in Japanese is a local anaphor (see (15a)), which, as we have seen, is not supported by a confirmed schematic asymmetry. In this section, I will try to provide further illustration of the proposed methodology in relation to lexical and structural hypotheses that are supported by confirmed schematic asymmetries in Japanese. Although it is abbreviated because of space considerations, the illustration to be given below is significant in light of the fairly common perception that it is not possible to obtain informant judgments that are of a categorical nature, which seems to lead to the view that empirical bases for our linguistic theorizing are bound to be gradient, probabilistic, etc.

Consider the example in (36).

(36) 55% izyoo-no tyuusyoo kigyoo-ga so-ko-no kaikeisi-o hihansita.

55%:or:more-GEN small:business-NOM that-place-GEN accountant-ACC criticized

'Each of 55% or more small businesses criticized its accountant.'

The structural condition for bound variable anaphora (henceforth BVA), such as the anaphoric relation holding between $55\%$ izyoo-no $N$ '55% or more Ns' and so-ko 'it' in examples like (36) is assumed to be universal. We adopt the thesis that the necessary condition for BVA(A, B), where $B$ is the intended dependent term and $A$ is its intended antecedent, is that what corresponds to $A$ c-commands what corresponds to $B$ at LF.50 Given this universal condition, we make a number of predictions about the availability of BVA(A, B) as long as we make independent hypotheses about the correspondences between phonetic strings in a particular language and LF representations—we shall refer to such correspondences as pf-LF correspondences.

The phonetic strings in question can be of the simple SOV form or of the OSV form; it can also be of a more complex form, such as what one might call "long-distance" OS form (so-called long-distance scrambling constructions) or "multiple OS" form (so-called multiple-scrambling constructions). The O of the OSV can be a direct object or an indirect object. We can also address cases of OSV where an overt element (such as so-ko 'it') appears in the canonical object position, which has been called "resumption" (in the OS construction) in Ueyama 1998, Hoji 2003 and the references there. In what follows, we will, however, only address the simple SOV.

The results of our experiments indicate, much in harmony with what is discussed in Hoji 2003 and 2006a, that it is indeed possible to obtain confirmed schematic asymmetries in regard to the availability of BVA(A, B). Consider, for example, the schematic representations in (37) and their somewhat more specific versions in (38).

49 The relevant point can be illustrated, for example, by making reference to the passives, the light verbs, and a "scope-bearing" predicate such as the potential --rare and the negative --(a)na(i) in Japanese. But space considerations prevent me from providing further discussion and illustration here.

50 This is in the spirit of Reinhart (1983), restating her proposal in terms of LF; see Reinhart 1983: 25, (19) and 26, (21).
(37) a. \( \text{ok Schema} \)
A-ga ... [ ... B ... ]-o ... V-T
BVA(A, B)

b. \*Schema
[ ... B ... ]-ga ... A-o ... V-T
BVA(A, B)

c. \( \text{ok Schema} \)
[ ... B ... ]-ga ... A-o ... V-T
B is referential.

(38) a. \( \text{ok Schema} \)
NP-ga ... [ ... so-NP ... ]-o ... V
BVA(NP, so-NP)

b. \*Schema
[ ... so-NP ... ]-ga ... NP-o ... V
BVA(NP, so-NP)

c. \( \text{ok Schema} \)
[ ... so-NP ... ]-ga ... NP-o ... V
(With so-NP "referring to" an individual/object that has been mentioned in the preceding discourse)

(38) is one of the sets of Schemata that have been covered in our on-line experiments. The examples in (39) are among the Examples we have used in the experiments.

(39) a. \( \text{ok Example} \) conforming to (38a):
55\% izyoo-no tihoozititai-ga so-ko-no syokuin-o sinbunsizyoo-de hihansita.
55\%:or:more-GEN local:government-NOM that-place-GEN employee-ACC newspaper-on criticized

'Each of 55\% or more local governments criticized its employees on a newspaper.'

b. \*Example conforming to (38b):
so-ko-no syokuin-ga 55\% izyoo-no tihoozititai-o sinbunsizyoo-de hihansita.
that-place-GEN employee-NOM 55\%:or:more-GEN local:government-ACC newspaper-ons criticized

'It's employees criticized each of 55\% or more local governments on a newspaper.'

c. \( \text{ok Example} \) conforming to (38c):
so-ko-no syokuin-ga 55\% izyoo-no tihoozititai-o sinbunsizyoo-de hihansita.
that-place-GEN employee-NOM 55\%:or:more-GEN local:government-ACC newspaper-ons criticized

'It's employees criticized each of 55\% or more local governments on a newspaper.'

As discussed in some depth in Hoji 2003 and 2006a, not every instance of BVA(A, B) is subject to the structural condition in question, and hence it is important to identify what choices of \textit{A} (and \textit{B}) give rise to a \textit{confirmed schematic asymmetry} in a simple case before proceeding to a more complex case (e.g., from SOV to OSV); cf. section 4.2. If we use 55\% izyoo-no \textit{N}'55\% or more \textit{Ns}' as the intended binder and \textit{so-ko 'it'} as the intended bindee, the average scores of the \textit{Schemata} in (38a), (38b), and (38c) are 97, 8, 96, respectively (with 22 informants (17 "linguistically informed" and 7 "linguistically naïve"\textsuperscript{51}), as of May 21, 2010). See section 4 of the

\textsuperscript{51} There is not much difference between the two types of informants, as indicated below, although the sample sizes may not be large enough to draw a more definitive conclusion.

<table>
<thead>
<tr>
<th>size</th>
<th>Schema A</th>
<th>Schema B</th>
<th>Schema C</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Language-3-5-3.doc
26/50
With *NP-sae 'even NP' as the intended binder and with so-ko 'it' as the intended bindee, the average scores of the *Schemata in (38a), (38b), and (38c) are 96, 5, 95, respectively (with 17 informants, as of June 20, 2010). The fact that the *Schemata in (38b) have gotten the scores around 5 is in sharp contrast with the result, reported in section 3.3, that the *Schemata under the lexical hypothesis in (15a) have gotten the scores over 50. If we use an NP such as subete-no N 'every N' and 3-nin-no N 'three Ns', the average score of the *Schemata is much higher (around 20), suggesting that BVA with such NPs as the intended binder need not be based on the LF c-command condition, as pointed out in Ueyama 1998. See section 4.6 of the supplementary file. This in turn indicates that the use of NPs such as 55% izyoo-no N '55% or more Ns' and NP-sae 'even NP' as A of BVA(A, B) makes experiments on OSV more directly revealing about the properties of the Computational System than the use of NPs such as subete-no N 'every N' and 3-nin-no N 'three Ns'.

If a *Schema-based prediction gets disconfirmed, we cannot tell a priori what hypothesis or hypotheses should be blamed for it; cf. the remarks on the Duhem thesis in section 4.2. In accordance with the "Maximize our chances of learning from errors" heuristic, it is therefore of utmost importance that we obtain a confirmed schematic asymmetry in a given experiment before proceeding to a more complex one where an additional or new hypothesis is to be tested. Otherwise, we would not be able to identify the culprit in the event that the *Schema-based prediction gets disconfirmed in the more complex experiment.

With regard to the various experiments alluded to above, it follows that experiments on OSV should be built on the proper "ingredients" (such as the appropriate choices of A and B in BVA(A, B)) that have given rise to a confirmed schematic asymmetry in experiments on SOV. The result of an OVS experiment would not tell us anything if it were designed crucially based on a hypothesis that has been shown in a prior SOV experiment not to be supported by a confirmed schematic asymmetry; see section 4.2.

5.2. Preliminary experiments

The informant judgment in experiments on the availability of BVA in SOV is revealing about properties of the Computational System only if the informant clearly understands what is meant by BVA. How can we then ensure, as much as possible, that the informants have a clear understanding of what is meant by BVA? In this subsection, we shall turn to experiments that are preliminary to those on the availability of BVA in SOV.

Japanese has three non-interrogative demonstrative prefixes ko- 'this', so- 'that', and a- 'that', as exemplified in (40) and (41).

(40) a. ko-no hito 'this person'
    b. so-no hito 'that person'
    c. a-no hito 'that person'

(41) a. ko-ko 'this place'
    b. so-ko 'that place'
    c. a-soko 'that place'

<table>
<thead>
<tr>
<th>linguistically informed</th>
<th>17 people</th>
<th>102 values</th>
<th>98</th>
<th>102 values</th>
<th>8</th>
<th>101 values</th>
<th>95</th>
</tr>
</thead>
<tbody>
<tr>
<td>linguistically naive</td>
<td>7 people</td>
<td>42 values</td>
<td>95</td>
<td>39 values</td>
<td>7</td>
<td>42 values</td>
<td>99</td>
</tr>
</tbody>
</table>

See Hoji 2006a for relevant empirical discussion.


The so in a-soko is unrelated to the demonstrative prefix so-.
Both so-NPs and a-NPs can be anaphorically related with another NP, as illustrated in (42).

(42) a. Toyota-ga soko-no sitauke-o hihansita.
   'Toyota criticized its subsidiaries.'
   Toyota-NOM that:place-GEN subsidiary-ACC criticized

b. Toyota-ga asoko-no sitauke-o hihansita.
   'Toyota criticized its subsidiaries.'
   Toyota-NOM that:place-GEN subsidiary-ACC criticized

It has, however, been observed in Nishigauchi 1986, Hoji 1990, Yoshimura 1992, and subsequent works that so-NPs can, but a-NPs cannot, be anaphorically related to a non-singular-denoting NP and the anaphoric relation in question has been considered as an instance of BVA. The generalization is recorded in (43) and is illustrated in (44).

(43) So-ko 'it, the/that place' and (to a lesser extent) so-itu 'the/that guy' can be anaphorically related to a non-singular-denoting expression, unlike a-soko 'it, the/that place' and a-itu 'the/that guy'.

(44) a. Kanarino kazu-no seizika-ga {so/*a}-itu/-no hisyo-o hihansita.
   considerable number-NOM politician-NOM that-guy-GEN secretary-ACC criticized
   'Each of a considerable number of politicians criticized that guy's secretary.'

b. 2wari izyoo-no zititai-ga [{so-ko/*a-soko}-o hihansita zassikisyaa]-ni
   20% more-GEN local:government-NOM that-place/that-place-ACC criticized magazine:reporter-DAT
   renraku-o totta.
   contact-ACC made
   'Each of 20% or more local governments made contact with a/the magazine reporter(s) who had criticized it.'

Leaving aside exactly how to express (43) in theoretical terms, the examples with a so-NP in (44) conform to the okSchema in (45a) and those with an a-NP in (44) to the *Schema in (45b).55

(45) (Where NP1 is not singular-denoting)
   a. okSchema:
      NP1-ga ... so-NP ... V
      BVA(NP1, so-NP)
   b. *Schema:
      NP1-ga ... a-NP ... V
      BVA(NP1, a-NP)
   c. okSchema:
      NP1-ga ... a-NP ... V
      (With a-NP "referring to" a particular individual/object.)

As of April 15, 2010, the average scores of the Schemata in (45a), (45b) and (45c) are 75, 5, 94, respectively, with 37 informants (22 "linguistically naïve" and 15 "linguistically informed"). See section 5 of the supplementary file. The *Schema-based prediction involving (45b) thus seems to have survived a disconfirmation attempt, in

55 There must be a language-specific lexical hypothesis that yields the effects of (43); Ueyama 1998 puts forth a formal proposal that has precisely those effects.
sharp contrast with the *Schema-based prediction based on (15a). 56

The claim that the anaphoric relation alluded to in (43) is an instance of BVA is strongly supported by the independent observation that expressions such as so-ko/so-itu are singular-denoting and hence the anaphoric relation in question cannot be regarded as an instance of coreference. We have conducted another set of preliminary experiments in which we test whether split antecedence is impossible with so-ko and so-itu, as pointed out in Hoji 1995: 2.2.

Consider the schematic representations of the Japanese examples in (46). So-itu-tati corresponds to 'those guys' or 'that guy and others'.

(46) a. *Schema:
NP1-ga NP2-ni ... so-itu-tati ... V
(so-itu-tati = NP1+NP2)

b. *Schema:
NP1-ga NP2-ni ... so-itu ... V
(so-itu = NP1+NP2)

c. *Schema:
NP1-ga NP2-ni ... so-itu ... V
(so-itu = NP1 or NP2)

The examples in (47) are on the basis of the Schemata in (46) and are among the examples included in our on-line experiment on the predicted schematic asymmetry in (46).

(47) a. *Example
Aru ninensei1-ga aru itinensei2-ni [sensei-ga so-itu-tati1,2-o hometeiru to]
certain sophomore-NOM certain freshman-DAT professor-NOM that-guy-tati-ACC was:praising that
tutae ta rasii.
told it:seems
'It seems that some sophomore1 told some freshman2 that the professor was praising them1,2.'

b. *Example
Aru ninensei1-ga aru itinensei2-ni [sensei-ga so-itu1,2-o hometeiru to]
certain sophomore-NOM certain freshman-DAT professor-NOM that-guy-ACC was:praising that
tutaeta rasii.
told it:seems
'It seems that some sophomore told some freshman that the professor was praising them.'

c. *Example
Aru ninensei1-ga aru itinensei-ni [sensei-ga so-itu-o hometeiru to]
certain sophomore-NOM certain freshman-DAT professor-NOM that-guy-ACC was:praising that
tutaeta rasii.
told it:seems
'It seems that some sophomore told some freshman that the professor was praising him.'

The examples in the experiment do not contain indices and the instructions given are like "under the interpretation that so-itu-tati 'those guys' refers to the sophomore and the freshman under discussion" (for (47a), for example),

56 This is a much more robust result than what is reported in Aoshima et al. 2009: 3.2 on the difference between a so-NP and an a-NP. I will return to this in sections 5.3 and 5.4.
"so-itu 'that guy' refers to the sophomore and the freshman under discussion" (for (47b), for example), so-itu refers to the freshman under discussion" (for (47c), for example).

As of May 26, 2010, the average scores of the Schemata in (46a), (46b), and (46c) are 76, 6, 81, respectively, with 26 informants. See section 6 of the supplementary file. Such results support the claim that so-ko and so-itu are singular-denoting, and hence confirm that it is reasonable to consider that the anaphoric relation intended in examples like (44) is an instance of BVA.

If some informants find *Examples of the form in (38b) acceptable, we should make sure that they have a clear understanding of BVA by checking their judgments on the *Examples conforming to (45b) and (46b), as compared to those conforming to (45a) and (46a), respectively. If we focus on the judgments by such informants when we consider the results in the experiments on the Schemata in (38b), the average score on the *Schemata becomes lower substantially, getting quite close to 0. This is illustrated in the supplementary file; see sections 9.5, 10.5, and 11.5. Notice that experiments on (45) and (46) concern lexical (and hence language-particular) properties and do not concern a structural condition for BVA while experiments on (38) are intended to test the validity of a universal structural hypothesis (as well as that of language-specific hypotheses about pf-LF correspondences). It is thus reasonable to consider the former types of experiments as preliminary to the latter in research concerned with properties of the Computational System of the language faculty.57 Section 8 of the supplementary file contains some relevant remarks thought stated only informally.

One might wonder what preliminary experiments could have been conducted so as to identify the reliable informants in experiments on (15a). If we can identify such "reliable informants" with regard to experiments on otagai, and if we focus on their judgments on examples in experiments on (15a), we might obtain a result that is different from what is reported above and the hypothesis in (15a) might be supported by the experimental results.

Recall that, if otagai were a local anaphor, examples like (48) should be completely unacceptable because there is no possible antecedent for otagai in examples like (48).

(48) Haru-no atatakana kaze-ga otagai-o totemo siawasena kimoti-ni sita.
    Spring-GEN warm wind-NOM otagai-ACC very happy feeling-DAT made

'The warm spring breeze made otagai (=them) feel very happy.'

The predicted unacceptability of examples like (48) is independent of any structural relation; it only has to do with the alleged lexical property of otagai. One may thus suggest that, if some informants accept examples like (48), at least to some extent, they should not qualify as reliable informants in an experiment on (15a), which is designed to test the effects of the structural relation between otagai and its intended antecedent.58 As it has turned out, however, there seems to be little hope for such an attempt to save (15a). As of May 24, 2010, none of the 19 informants (14 "linguistically informed" and 5 "linguistically naïve"—see footnote 30) finds examples like (48) completely unacceptable. The lowest score is 33 by a "linguistically informed" informant and the next lowest score is 72 by a "linguistically naïve" informant, again on the 0-100 scale. The majority of the informants find examples like (48) fully acceptable. The average scores of the *Schema in question are 93 for the "linguistically informed" informants and as well as for the "linguistically naïve" informants. See section 7 of the supplementary file.

The illustration given above of how confirmed schematic asymmetries obtain in Japanese is rather sketchy due to the space limit and a fuller illustration and discussion of various related issues—including the formal

57 Subsequent to the completion of the draft of this paper in August of 2010, we have obtained experimental results, with two different groups of undergraduate students, at two different institutions, with no prior linguistics background, that provide a more compelling case for the point under discussion, with regard to the significance of paying attention to the correlation of informant judgments between experiments (as well as the effectiveness of certain binders). In the interest of space, I do not include the relevant discussion here, however.

58 One may, for example, suggest that those informants do not have the lexical specification for otagai under discussion; but see footnote 74.
5.3. Effects of the "instructions" in an experiment

Consider again the predicted schematic asymmetry in (45) discussed in section 5.

\[(45) \quad \text{(Where } NP1 \text{ is not singular-denoting)}
\]
\[\text{a. } \text{okSchema:}
\quad NP1...so-NP V
\quad BVA(NP1, so-NP)
\]
\[\text{b. } \text{*Schema:}
\quad NP1...a-NP V
\quad BVA(NP1, a-NP)
\]
\[\text{c. } \text{okSchema:}
\quad NP1...a-NP V
\quad \text{(With } a\text{-NP "referring to" a particular individual/object.)}
\]

\[(45b) \quad \text{are a } \text{*Schema because of the hypothesized lexical property of } a\text{-NPs that is responsible for the generalization in (43).}
\]

\[(43) \quad \text{So-ko 'it, the/that place' and (to a lesser extent) so-itu 'the/that guy' can be anaphorically related to a non-singular-denoting expression, unlike } a\text{-soko 'it, the/that place' and } a\text{-itu 'the/that guy'.}
\]

As noted in section 5, the mean score on (45b) is 5 on the 0-100 scale.

Aoshima, Yoshida and Phillips (2009: 3.2) (henceforth AYP) report the result of an experiment, which, like our experiment on (45), tests the predicted difference between a so-NP and an a-NP used as the "intended bindee," i.e., with the relevant NPs being used as B of BVA(A, B) in the terms of the preceding discussion. The *Schemata used in their experiment are as in (49).

\[(49) \quad \text{Aoshima et al.}'s *Schemata:}
\]
\[\text{a. } [... a\text{-NP } ... ]\text{-ga } NP1 V
\quad BVA(NP1, a\text{-NP})
\]
\[\text{b. } [... a\text{-NP } ... ]\text{-o } NP1 V
\quad BVA(NP1, a\text{-NP})
\]

Further illustration of confirmed schematic asymmetries in Japanese can be provided on the basis of the results of more involved experiments, including those on the availability of BVA in the OSV (including paradigms of "long-distance" OSV and also with "resumption") and those concerning local disjointness effects. Such results as well as results of experiments that deal with quantifier scope, "floating numerals," negation, passives, etc. will be provided in separate works.

Such a practice, I should like to add, is by no means limited to (15a) (and also (15b) and (15c)), as one can confirm by studying published works dealing with hypotheses concerning BVA and quantifier scope, for example. See footnote 44.
Given the hypothesized lexical property of \(a\)-NPs that is responsible for the generalization in (43), it is predicted that examples confirming to (49a) or (49b) are completely unacceptable under the intended BVA construal.\(^{61}\) The mean score on AYP's *Examples conforming to (49a) and (49b) are reported to be 1.71 and 1.95 on the scale of 1-5, respectively, (with 28 informants (all undergraduate students)), contrasting sharply with the mean score of 5 on (45b) (on the 0-100 scale) in our experiment. See section 2.2 of the supplementary file. As discussed in section 5, the mean scores of the Schemata in (45a), (45b) and (45c) are 75, 5, 94, respectively, with 37 informants (22 "linguistically naïve" and 15 "linguistically informed"), as of April 15, 2010. Once we convert AYP's scores to the 0-100 scale, we obtain the following mean scores on (45b), (49a) and (49b)

\[\text{(50)} \quad \text{The mean scores on (45b) in our experiment and (49a) and (49b) in AYP's:}\]

- a. (45b): 5
- b. (49a): 21.375
- c. (49b): 24.375

Although AYP (2009: 3.2) concludes that "the large difference in acceptability ratings between \textit{soko} and \textit{asoko} lends support to the claim by Ueyama and colleagues that \textit{soko} allows bound-variable readings in Japanese whereas \textit{asoko} does not," we would not consider a mean score of 20-25 (on the 0-100 scale) on a *Schema as support for the lexical hypothesis that gives rise (43). We would instead consider such a mean score as a disconfirmation of the *Schema-based prediction, in accordance with the methodology suggested above.\(^{62, 63}\)

There is a difference between (45b) and (49); the \(a\)-NP is preceded by its intended antecedent in the *Schema in (45b), but not in those in (49). It is, however, highly unlikely that this is the source of the difference in informant judgment because the relevant lexical property of an \(a\)-NP should result in the impossibility of BVA(NP, \(a\)-NP) regardless of the structural relation between the \(a\)-NP and its intended antecedent.

There is another difference between our on-line experiments (as reported in section 5) and AYP's experiment; the intended binders in our experiments are \textit{kanarino kazu-no N} 'a considerable number of Ns' and \textit{2wari izyoo-no N} '20\% or more Ns' while the intended binder in AYP's is \textit{dono N-mo} 'whichever NP/every NP'. One might thus wonder whether the difference in informant judgment in question is due to, or at least related to, the choice of the intended binder. As pointed out briefly in section 5 and as will be further discussed below, the choice of A of BVA(A, B) does affect informant judgments on the availability of BVA(A, B). There is, however, reason to believe that the difference between the mean score on (45a) in our experiment and those on (49) in AYP's cannot be attributed to the choice of A of BVA(A, B). In one of the experiments we have conducted but have not discussed in this paper, we use \textit{do-no N-mo} 'whichever NP/every NP' as the intended binder in the Examples conforming to the schemata in (45). As of July 11, 2010, the mean scores for Schemas A, B, and C in (45) in that experiment are 82, 7, 85, respectively, with 48 informants (43 "linguistically naïve" and 5 "linguistically informed"). See section 12.2 of the supplementary file. These results are fairly close to those of the experiment

\[\text{61 AYP is concerned with "sentence processing." One may say that AYP's main concern is what goes on when a numeration is formed, in the terms of the diagram in (4), leaving aside the question of whether or not AYP would in fact adopt (4). Given the lexical hypothesis under discussion (that has the effect of (43)), which AYP also seem to adopt, the predicted informant judgment on examples conforming to (49) is 0, nonetheless.}\]

\[\text{62 See footnote 23 for related remarks.}\]

\[\text{63 Given in (i) are their Schemata with a \textit{so}-NP as the intended bindee, which they considered along with (49).}\]

(i) a. \[[... \textit{so}-NP ...] \text{-} \text{ga} \text{ NP1 V}\]
   \text{BVA(NP1, so-NP)}

b. \[[... \textit{so}-NP ...] \text{-} \text{o} \text{ NP1 V}\]
   \text{BVA(NP1, so-NP)}

According to AYP, the mean scores on (i-a) and (i-b) are 3.21 and 2.49 (see their Table 2)—which would correspond to 55.25 and 37.25, respectively, on the 0-100 scale—respectively.
on (45) noted above where the binders are *kanarino kazu-no N* 'a considerable number of Ns' and *2wari izyoo-no N* '20% or more Ns', with the mean scores on Schemata A, B, and C being 75, 5, and 94, respectively, as noted above. The mean score of 7 on (45b) with *dono N-mo 'whichever NP/every NP'* in our experiment thus suggests that the use of *dono N-mo 'whichever NP/every NP'* is most likely not responsible for the mean score being higher than 20 on the *Schemata* in (49a) and (49b) in AYP's experiment. What could then be the source of the difference between the mean score on (45b) with *dono N-mo 'whichever NP/every NP'* as the 'intended binder' in our experiment and those on (49a) and (49b) in AYP's, in which the 'intended binder' is also *dono N-mo 'whichever NP/every NP'? Moving to the "word order," AYP's experiment and ours with *dono N-mo 'whichever NP/every NP'* as the intended binder differ from each other with regard to how the 'instructions' are given to the informants. According to AYP 2009: 115, their informants "were asked to judge whether an underlined pronoun and an underlined quantificational NP in each sentence could be understood as coreferential, indicating their rating on a scale from 1 (coreference impossible) to 5 (coreference fully acceptable)." Leaving aside exactly how their "instructions" were given in Japanese, it is possible that some of the informants did not understand what was meant by "coreferential." If some of the informants do not clearly understand what is meant by the intended interpretation, the results of an experiment cannot be very revealing about the properties of the Computational System under investigation, and it is not surprising that we obtain a result rather different from the predictions.

We conjecture that the mean scores on AYP's (49a) and (49b) are substantially higher than 0 (see (50)64) because the informants do not clearly understand the "instructions" in the experiment. In our on-line experiment under discussion, on the other hand, the "instructions" for each sentence (see section 3.3) are given as in (51) when translated into English.65

(51) Can you express the situation described in the parentheses by the sentence below? If you cannot express (the described situation) (with the sentence below) no matter what, please choose "x." If you can express (the described situation) (with the sentence below), please choose "o." If you are not sure, please proceed (to the next question) without answering.

The "situation" is given as in (52) for the sentences in (53a) and (53b), and as in (54) for the sentence in (55), translated here into English.

(52) Instructions for (53a) and (53b): (When Hansin, Hirosima, Yokohama, Tyuuniti, Kyozin, and Yakuruto are being talked about: the situation where Hansin criticized Hansin's batting coach, Hirosima criticized Hirosima's batting coach, Yokohama criticized Yokohama's batting coach, Tyuuniti criticized Tyuuniti's batting coach, Kyozin criticized Kyozin's batting coach, and Yakuruto criticized Yakuruto's batting coach.)66

(53) a. *Example conforming to (45a):* 
 Dono kyuudan-mo so-ko-no dageki kooti-o hihansita.  
 which ball-club ALSO that-place-GEN hitting coach-ACC criticized

'Every ball club criticized its hitting coach.'
b. *Example conforming to (45b):

Dono kyuudan-mo a-soko-no dageki kooti-o hihansita.
which ball:club- ALSO that-place- GEN hitting coach-ACC criticized

'Every ball club criticized its hitting coach.'

(54) Instructions for (53a) and (53b):
(When Hansin, Hirosima, Yokohama, Tyuuniti, Kyozin, and Yakuruto are being talked about: the situation where Hansin, Hirosima, Yokohama, Tyuuniti, Kyozin, and Yakuruto criticized Tyuuniti 's batting coach.)

(55) Example conforming to (45c):

Dono kyuudan-mo a-soko-no dageki kooti-o hihansita.
which ball:club- ALSO that-place- GEN hitting coach-ACC criticized

'Every ball club criticized its hitting coach.'

Assuming that the instructions as given in our on-line experiment under discussion make the "intended interpretation" clearer to the informants than those given in AYP's experiment, we can attribute the difference between the mean scores on AYP's (49a) and (49b) and that on our (45a) to how the "instructions" were given to the informants in the two experiments.67

5.4. Effects of the choice of the intended binder (A of BVA(A, B))

As noted in footnote 62, AYP's (2009: 3.2) Schemata with a so-NP as the intended bindhee are as in (56).

(56) AYP's (2009: 3.2) Schemata with a so-NP

a. [...] so-NP ...]-ga NP1 V
BVA(NP1, so-NP)
b. [...] so-NP ...]-o NP1 V
BVA(NP1, so-NP)

According to Aoshima et al., the mean scores on (56a) and (56b) are 3.21 and 2.49 on the 1-5 scale (see their Table 2), respectively, which would correspond to 55.25 and 37.25, respectively, on the 0-100 scale, as summarized in (57). See section 2.2 of the supplementary file.

(57) The mean scores on (56a) and (56b) in AYP's experiment, on the 0-100 scale:

a. (56a): 55.25
b. (56b): 37.25

What is particularly noteworthy is the mean score of 55.25 on (56a).68 In section 5, we discussed the schemata in (38), which I repeat here.

(38) a. *Schema
NP-ga ...[ ... so-NP ...]-o ... V
BVA(NP, so-NP)

67 We have further reason for this conjecture, on the basis of substantially different informant judgments on the same set of examples, depending upon how the "instructions" are given to the informants. We plan to report on that in a separate work. See section 12 of the supplementary file.

68 It is also noteworthy that the mean score on (56a) is higher than (56b). I will return to this later.
b. *Schema
    [ ... so-NP ...]-ga ... NP-o ... V
    BVA(NP, so-NP)

c. okSchema
    [ ... so-NP ...]-ga ... NP-o ... V
    (With so-NP "referring to" an individual/object that has been mentioned in the preceding discourse)

The schemata in (38b) and (56b) are a typical configuration of so-called Weak Crossover. In the terms of the present discussion, examples conforming to (38b) and those confirming to (56a) are predicted to be completely unacceptable under BVA(NP, so-NP) as long as (i) the BVA(NP, so-NP) in question is hypothesized to arise only if there is a structural relation of c-command at LF holding between what corresponds to NP (the intended binder) and what corresponds to so-NP (the intended bindee) at LF—more specifically, the former c-commanding the latter at LF—and (ii) it is hypothesized that the object (NP1 in (56a) and the NP-o in (38b)) fails to c-command the subject, hence what is contained in it, including the so-NP in those schemata. Among the Examples we have used in our on-line experiments on (38) are the examples in (39), repeated here.

(39) a. okExample conforming to (38a):
    55% izyoo-no tihoozititai-ga so-ko-no syokuin-o sinbunsizyoo-de hihansita.
    55%:or:more-GEN local:government:NOM that-place-GEN employee-ACC newspaper-on criticized
    'Each of 55% or more local governments criticized its employees on a newspaper.'

b. *Example conforming to (38b):
    so-ko-no syokuin-ga 55% izyoo-no tihoozititai-o sinbunsizyoo-de hihansita.
    that-place-GEN employee-NOM 55%:or:more-GEN local:government-ACC newspaper-on criticized
    'Its employees criticized each of 55% or more local governments on a newspaper.'

c. okExample conforming to (38c):
    so-ko-no syokuin-ga 55% izyoo-no tihoozititai-o sinbunsizyoo-de hihansita.
    that-place-GEN employee-NOM 55%:or:more-GEN local:government-ACC newspaper-on criticized
    'Its employees criticized each of 55% or more local governments on a newspaper.'

The English translations of the "intended interpretation" given to the informant for (39a), (39b), and (39c) are as in (58a), (58b), and (58c), respectively.

(58) a. Under the rendooyomi (co-variant reading) between 55% izyoo-no NP and soko; i.e., under the interpretation that the local governments that criticized their own (zibun-no tokoro-no 'self-GEN place-GEN) employees on a newspaper are 55% or more of the entire local governments.

b. Under the rendooyomi (co-variant reading) between 55% izyoo-no NP and soko; i.e., under the interpretation that the local governments that were criticized by their own (zibun-no tokoro-no 'self-GEN place-GEN) employees on a newspaper are 55% or more of the entire local governments.

c. Under the interpretation that soko refers to 'Department of Finance' that has been mentioned in the preceding discourse.

As noted, the predictions are that *Examples conforming to (38b), such as (39b), are completely unacceptable,
as long as BVA(A, B) is hypothesized to arise only if there is the c-command relation holding at LF between the LF objects corresponding to A and B. That is to say, complete unacceptability is predicted for *Examples such as (39b) in part because of the researcher's commitment to the hypothesis that BVA(55% izyoo-no N, so-NP) is possible only if there is the c-command relation at LF holding between the LF objects corresponding to 55% izyoo-no N '55% or more Ns' and so-NP 'the/that NP'. As noted in section 5, the mean score on the *Examples conforming to (38b), with the intended binder being 55% izyoo-no N '55% or more Ns', such as (39b), was 8. The fact that it is substantially higher than 0 calls for an explanation; but we can reasonably hope that we might be able to obtain a mean score closer to 0 as we come to be able to identify the noise and as we improve on the experimental design and interpretations. See section 9.5 of the supplementary file.

The mean score on (56a) in AYP's experiment, on the other hand, is over 50, contrary to the predicted 0. One might attribute the difference between the mean score on our (38b) and that on AYK's (56a) to the instructions given to the informants, as in the case of the disconfirmation of the *Schema-based prediction concerning (49), and hence to the informants' insufficient understanding of the relevant interpretation in question. But that cannot seem to be the only reason.

Recall that the predictions in question are that *Examples conforming to (38b), such as (39b), are completely unacceptable, as long as BVA(A, B) is hypothesized to arise only if there is the c-command relation holding at LF between the LF objects corresponding to A and B. As discussed in some depth in Hoji 2003 and 2006a, and as briefly noted in section 5, not every instance of BVA(A, B) is clearly subject to the structural condition in question. For example, as also mentioned in section 5, with 55% izyoo-no N '55% or more Ns' as the intended binder and so-ko 'it' as the intended bindee, the average scores of the Schemata in (38a), (38b), and (38c) are 97, 8, 96, respectively (with 22 informants, as of May 21, 2010). With NP-sae 'even NP' as the intended binder and with so-ko 'it' as the intended bindee, the average scores of the Schemata in (38a), (38b), and (38c) are 96, 5, 95, respectively (with 17 informants, as of June 20, 2010). If we use an NP such as subete-no N 'every N' and 3-nin-no N 'three Ns', however, the average score of the *Schemata is much higher (around 20), suggesting that, as pointed out in Ueyama 1998, BVA with such NPs as the intended binder need not be based on the LF c-command condition.

A natural question to ask is thus whether or not BVA(dono-N-mo, so-ko) arises only on the basis of the c-command relation in question. We have not yet conducted experiments to answer the question in accordance with what has been suggested in the main text. We did, however, conduct experiments, starting several years ago, where we tested the sensitivity of the availability of BVA(A, B) to LF c-command, with a number of choices of A and B of BVA(A, B). In one of those experiments, the following schemata are considered.

(59) (Under BVA(A, B))

a. A-ga [... B ... ]-o V-T
b. [... B ... ]-ga A-o V-T
c. [... B ... ]-o A-ga V-T
d. A-o [... B ... ]-ga V-T
e. [... A ... ]-ga B-o V-T

In that experiment, while B of BVA(A, B) remains to be so-ko 'that place, it', 9 choices are tried for A of BVA(A, B), including 10 izyoo-no N '10% or more Ns', subete-no N 'every N', A to B to 'both A and B', NP-sae 'even NP', NP-dake 'only NP', dono N 'which N', and dono-N-mo 'whichever N, every N', with 5x9=45 examples altogether. The results of the experiment that are of the immediate relevance here are: with 10 izyoo-no N '10% or more Ns' as A of BVA(A, B), we obtain fairly robust informant judgments, but with dono-N-mo, 'whichever N, every N', we do not. That is to say, with the same instructions given to the informants, there is a significant difference between the mean score on the schema in (59b), depending upon whether the A of BVA(A, B) is 10 izyoo-no N '10% or more Ns' or dono-N-mo 'whichever N, every N'. See 4.6 of the supplementary file. It seems clear that the mean score of 55.25 on the *Schema in (56a) cannot be attributed solely to how the instructions are given to the informants. It seems that even if the informants understand clearly what is intended by BVA(dono-N-mo, soko), the informants' judgments on examples conforming to (59b), with dono-N-mo 'whichever N, every N' as the intended binder, are most likely much higher than 0.
Further considerations confirm that we can attain a mean score on a *Schema that is closer to 0 only if we have improved our experimental design, including how we give the instructions to the informants, and how we "interpret" the results of an experiment in accordance with the "Maximize our chances of learning from errors" heuristic, by taking into account the results of preliminary experiments, not only in terms of the effects of the choice of various relevant lexical items but also in terms of how the informants have responded in preliminary experiments. Elaboration of the point would have to refer to the details of results of a number of experiments and it will have to be made in a separate paper, due to space considerations.

6. Some implications

6.1. On the use of statistics

Reliance on statistics is rather minimal under the proposed methodology of hypothesis testing in research concerned with properties of the Computational System. Its most crucial aspect is whether we obtain a confirmed schematic asymmetry, and that will be contingent upon whether any *Example conforming to any *Schema is judged by every informant to be completely unacceptable. There may be room for the use of statistics in regard to what is to count as "complete unacceptability" for a given *Example, in light of the fact that informants might make an error in reporting their judgments. But such use of statistics should be understood to be quite distinct from the use of statistics in research that heavily relies on 'statistically significant contrasts'.

Under the proposed method of evaluating hypotheses, a contrast in acceptability is significant only if a *Schema-based prediction survives a rigorous disconfirmation attempt. A 'statistically significant contrast' such as one consisting of 30 on a *Schema and 60 (or even 100, for that matter) on the corresponding *Schemata, on the 0-100 scale, does not count as a confirmed schematic asymmetry according to the proposal suggested here. The thesis we adopt that what is considered to be a likely reflection of properties of the Computational System is a zero vs. non-zero contrast, i.e., a contrast between complete unacceptability and the lack thereof, rather than a 'statistically significant contrast' separates the present approach, in a crucial way, from most other "experimental approaches" in the field, including Cowart 1997 and works in the Magnitude Estimation approach; cf. Sprouse 2007: 124 and Young 2008: 209-10 for some relevant discussion.

Nakaya's 1958: 17 characterization of science that the reproducible phenomena (that we extract from nature) must be analyzable statistically is in line with a commonly-held view; it is stated in Weinberg 1992:7 as: "[w]hat a successful scientific explanation would have to accomplish" is "the quantitative understanding of phenomena." (The emphasis is as in the original.) One may speculate that heavy reliance on 'statistically significant contrasts'

70 Morrison and Henkel 1970/2007 contains a number of works that critically address the use of significance tests in social sciences (outside linguistics), including Meehl 1967: 252, which states as a "paradox," that "[i]n the physical sciences, the usual result of an improvement in experimental design, instrumentation, or numerical mass of data, is to increase the difficulty of the observational hurdle" which the physical theory of interest must successfully surmount; whereas, in psychology and some of the allied behavioral sciences, the usual effect of such improvement in experimental precision is to provide an easier hurdle for the theory to surmount." Meehl elaborates on the "puzzle" and observes (p. 264), "In physics, the substantive theory predicts a point-value, and when physicists employ "significance tests," their mode of employment is to compare the theoretically predicted value $x_0$ with the observed mean $\overline{x}_0$, asking whether they differ (in either direction!) by more than the "probable error" of determination of the latter," contrasting it with fields where predictions are only on a difference or a tendency. In such fields, the testing of a hypotheses is indirect; what is subjected to refutation, so to speak, is the "null hypothesis" that there is no such difference/tendency, rather than the substantive hypotheses that entail the existence of such a difference/tendency; if the probability of the observed difference/tendency is less than a certain (arbitrarily chosen) threshold (often 5%, but sometimes 1% or 10%), the null hypothesis gets rejected, which in turn is taken as support for the substantive hypotheses. This is the most crucial aspect of significance tests in social, behavioral and life sciences, as I understand it. Meehl's point is thus that the more measurement precision we attain, the more likely the hypotheses that are "directly" tested get rejected; in the case of physics, they are the hypotheses under discussion whereas in social and behavioral sciences, they are the null hypotheses. The more likely the null hypothesis gets rejected, therefore, the more likely the substantive hypotheses get "corroborated" in the latter fields. Cf. also Lakatos 1970: 176, note 1 (reproduced in Lakatos 1978: 88-89, note 4) for relevant remarks.
in disciplines outside physical sciences, such as linguistics, stems from adopting this focus on "the quantitative understanding of phenomena" and at the same time realizing the immense difficulty in obtaining data of a categorical nature that are deducible from a proposed system; see Meehl 1967 for much relevant discussion. One of the main theses of the present paper is that it is possible to deepen our understanding of the properties of the Computational System by rigorously applying the hypothetico-deductive method, on the basis of data that are of a categorical nature, namely, confirmed schematic asymmetries. To the extent that we can achieve that, we will have shown that "the quantitative understanding" of the properties of the language faculty (more precisely, those aspects of the language faculty that can be studied "scientifically") is in fact in terms of zero vs. non-zero contrasts. We will thus have shown that it is possible to make point-value predictions in language faculty science, and that should have far-reaching implications for research on the human mind beyond the language faculty per se, and also for what one can aspire to achieve in fields outside physics and closely related fields.

6.2. On cross-linguistic research

The fundamental challenge in trying to ensure the testability of hypotheses about the Computational System stems from the fact that, while hypotheses about the Computational System are universal and are necessarily quite abstract, actual experiments to test the validity of those hypotheses must be concrete because they must deal with specific predictions in a particular language. For example, although the empirical consequences of hypotheses that have been backed up by a confirmed schematic asymmetry in one language should in principle be testable with respect to any other language, lexical differences among languages tend to make it difficult, if not overwhelming, to replicate the same confirmed schematic asymmetry among different languages.

For this reason, the designs of experiments to test the same hypotheses about the Computational System may look quite different on the surface, depending upon what language is dealt with in a particular experiment. What is crucially needed is thus a clear articulation of the correspondences between 'sentence forms' in a particular language and the abstract representations that they correspond to—what has been referred to above as pf-LF correspondences—, along with the articulation of the relevant hypotheses about the Computational System, (and some lexical items), and the relevant bridging statement. And that would require a great deal of rigorous work of establishing confirmed schematic asymmetries in each of the languages under consideration. While meaningful comparison among languages can thus be carried out only at some level of abstraction, it must be recalled that no matter how abstract our theory of the language faculty may become, its empirical consequences should remain expressible, ultimately, in terms of confirmed schematic asymmetries.

6.3. Anomalies and the failure to obtain a confirmed schematic asymmetry

Being faced with the disconfirmation of the *Schema–based predictions under (15a), one might make recourse to the notion that science progresses in the ocean of anomalies anyway (see Lakatos 1970: 48-52 and Feyerabend 1975: chapter 5, for example), and maintain that the methodological proposal made in the preceding discussion goes against what is practiced in mature sciences; cf. Boeckx 2006: 89, 91, for example, for remarks that one might regard as endorsing such a view. I should like to note that anomalies in mature sciences, such as physics, are solid and in fact very precise observations that resist an account within a given research program, e.g., the anomalous orbit of Uranus (before the discovery of Neptune) and the anomalous precession of Mercury's orbit, to mention two of the most celebrated instances of anomalies within the Newtonian research program. What we have discussed above is the failure to obtain a confirmed schematic asymmetry, which I maintain should be regarded as a "basic unit of facts" for research concerned with the properties of the Computational System. The failure to obtain a confirmed schematic asymmetry should not be likened to an anomaly in mature sciences. After

\footnote{For example, the value of the anomalous precession of Mercury's orbit was 43 seconds per century. That is the discrepancy between what is predicted under the Newtonian theory and the actual observation. Unlike the case of anomaly with the orbit of Uranus, which had led to the discovery of Neptune, the Mercury anomaly resisted an explanation under the Newtonian theory and was eventually explained by, and provided the first empirical support for, Einstein's general theory of relativity. The precision of measurement that compels researchers in quantum physics to reconsider their hypotheses is even more staggering and almost mind boggling.}
all, in my assessment, we are still at a stage where we are trying to identify "basic units of facts" for research that approaches the Computational System of the language faculty by the hypothetico-deductive method. We will face an anomaly only after we have obtained a sufficiently large number of confirmed schematic asymmetries; an anomaly arises when some confirmed schematic asymmetries appear to resist a coherent account with respect to the rest of the confirmed schematic asymmetries.\footnote{One may even suggest, somewhat paradoxically, that we are at a stage where we are trying to accumulate confirmed schematic asymmetries so as to be able to identify an anomaly.}

7. Concluding remarks

Adopting the general model of the Computational System in Chomsky 1993, a model of judgment-making suggested in Ueyama 2010, and a research heuristic "Maximize our chances of learning from errors," I have suggested a means to extract informant judgments that are likely a reflection of properties of the Computational System.\footnote{I assume that properties of the language faculty that fall outside the Computational System can be most effectively studied in line with the "Maximize our chances of learning from errors" heuristic if we build our research on the results we will have obtained with regard to the properties of the Computational System.} The main claim of the present work is that our hypotheses about the Computational System of the language faculty are to be tested on the basis of data of a categorical nature, i.e., in terms of the contrast between complete unacceptability and the lack thereof. Not only is that necessary for the purpose of making our hypotheses testable in accordance with the hypothetico-deductive method, but it is also a consequence of adopting the model of judgment-making in (4) and postulating a bridging statement that specifies a necessary condition for a particular interpretation involving two linguistic expressions $a$ and $b$, $\gamma(a, b)$; see section 2.2, including footnote 9.

More specifically, I have argued that our empirical research must be built on confirmed schematic asymmetries, which I suggest should be regarded as "basic units of facts" for research that aims at a discovery of the properties of the Computational System of the language faculty by the hypothetico-deductive method. According to the proposal, a confirmed schematic asymmetry obtains if and only if informants' judgments on $\textit{Examples}$ are consistently "completely unacceptable" and their judgments on the corresponding $\textit{okExamples}$ are "acceptable" (at least, to some extent); see section 2.3. Thus, even if there are some speakers who detect a significant contrast among some relevant examples in question, that in and of itself is not of much significance.\footnote{If one wishes to 'save' a given language-specific hypothesis by making recourse to a dialect (or an idiolect), one must show (i) that (a) confirmed schematic asymmetry/ies indeed obtain(s) for the speaker(s) under discussion and (ii) what new $\textit{Schema}$-based prediction(s) can be made and be tested with that/those speaker(s). One should also specify some plausible way in which the relevant lexical property has been acquired by that/those speaker(s) but not by the others, given the assumption that the Computational System is invariant, with the possible exception of the so-called head parameter, and hence what is responsible for the dialectal/idiolectal difference in question must be of a lexical nature.}

One might wonder if this is an unrealistically high standard for actual research because we cannot fully control various non-grammatical factors. In response to such a possible objection, I have provided a brief demonstration in section 5 that it in fact seems possible to obtain a confirmed schematic asymmetry in accordance with the above-mentioned standard.\footnote{In separate works, we plan to provide further illustration on the basis of experiments that deal with the OSV (Object Subject Verb) order (with and without "resumption"), the local disjointness effects of Binding Principle B, quantifier scope, Negation, among other things.} The demonstration is significant especially in light of the central role played by introspective judgments in generative grammar. Despite the wide acceptance of the thesis that a major source of evidence for or against our hypotheses is informant judgments, the field has so far failed to articulate in what way informant judgments can be revealing about the properties of the Computational System, as can be seen,
for example, from the general absence of a model of judgment-making in the literature; cf. footnote 10. Objections have thus been raised over the years against the use of introspective judgments itself, and especially, the use of researcher's own judgments. Combined with frequently observed judgmental fluctuation, variation, and disputes, this seems to have given rise to a general perception that the study of language faculty cannot be an "exact science." 78

The present work maintains that data for language faculty science are confirmed schematic asymmetries, which are reproducible phenomena that are "measurable," in the terms of Nakaya's characterization of science; see section 4.1 above. Notice that it is not clear how non-categorical judgments can be "measured," without introducing some arbitrary criteria. As noted, I maintain that, if an alleged linguistic generalization is not supported by, or does not constitute, a confirmed schematic asymmetry, it is not (yet) part of data for language faculty science although it might well be part of a study of language.

There are a number of merits to making crucial reference to confirmed schematic asymmetries as "basic units of facts" against which we evaluate our hypotheses about the Computational System of the language faculty. For example, technical details tend to make things opaque unless conscious efforts are made to articulate how the proposal under discussion can be put to empirical test. Crucial reliance on confirmed schematic asymmetries helps us understand actual empirical consequences of various proposals, beyond the technical details and differences of frameworks, and makes it much more straightforward to determine which of the alternative proposals are to be preferred over the others in terms of their empirical merit (without relying on rhetorical skills). It also helps us understand how we can deal with, and/or proceed in, cross-linguistic research in a meaningful and effective way. If it becomes the norm in the field to evaluate an alleged generalization in a given language based on whether it is backed up by a confirmed schematic asymmetry, that will greatly enhance the reliability of generalizations reported about any language. Assessing the empirical merit of our hypotheses about the Computational System on the basis of confirmed schematic asymmetries should make discussion and debates with regard to the "scientific status" of generative grammar, the nature of data in generative grammar, etc. much more empirically grounded; see NLLT in 2000-2001 (including Lappin et al. 2000a, 2000b, 2001 and the responses to Lappin et al. 2000a and 2000b in the relevant issues), Devitt 2006 and Fitzgerald 2010 in British Journal of Philosophy of Science, the 2005 special issue of Lingua on "Data in Theoretical Linguistics," the 2007 special

76 It is perhaps safe to state that until the introduction of the model of the Computational System in Chomsky 1993, it was not possible to embed the model of Computational System within a model of judgment-making in any straightforward way so as to enhance the testability of hypotheses about the Computational System, which in turn makes it unclear what significance could be assigned to the actual informant judgments we obtain. I should like to conjecture that the lack of a model of judgment-making such as the one adopted here has resulted in the failure of the field at large to come to terms with what has been identified in the preceding discussion as the fundamental asymmetry between "Schema-based predictions and "Schema-based predictions, which, I should further like to suggest, has in turn often resulted in a confusing state of affairs when being faced with judgmental dispute on a crucial paradigm. Heavy reliance on "statistically significant contrasts" in "experimental" works in recent years seems to me to be a result of this. Recall that, according to the methodology proposed above, a "statistically significant contrast," in and of itself, by no means guarantees that the informant judgments under discussion may be a reflection of properties of the Computational System unless the *Schema-based prediction has survived a rigorous disconfirmation attempt.

77 See Schütze 1996 and the references there, for example; cf. also the other works cited in footnote 1.

78 Remarks such as "My personal experience, sad to say, is that it is difficult to convince my colleagues in philosophy and the physical sciences that grammatical theory in ANY shape or form is—or has the potential to be—scientific. And nothing leads them to tune out faster than to hear grammatical theory compared to physical theory," found in the last paragraph of section 1 of Newmeyer 2008, are thus not unexpected. In the paragraph containing the above remark, Newmeyer states "I find B [=Boeckx]'s extensive appeals to higher scientific authority to be quite tedious and I harbor the suspicion that advocates of any theory of language imaginative could find quote after quote from Galileo, Newton, or Darwin to help justify their approach." The broader point made by Newmeyer (2008) in that paragraph raises various issues that deserve discussion, including the relation between language faculty science and grammatical theory, among other things; but there is no space for the discussion here.
issue of *Theoretical Linguistics* volume on "Data in generative grammar," the 2002 paper by H. Lasnik and the related papers in 2003 in *TRENDS in Cognitive Sciences*, and Newmeyer 2008, among many other places. It is hoped that confirmed schematic asymmetries will serve a role very much like the observations made by Tycho Brahe, as suggested in section 2.4. Most importantly, crucial reference to confirmed schematic asymmetries as "basic units of facts" against which we evaluate our hypotheses about the Computational System of the language faculty makes us hopeful that we might be able to make generative grammar an empirical science, or to put it more accurately, to make language faculty science possible, where the general scientific method in (1) can be rigorously applied.79

8. References


79 One may take this statement as contentious if one thinks that generative grammar has already established itself as a science of the language faculty. As suggested in the preceding pages, my own assessment is rather different from such a view as long as we mean by "science" a field in which a hypothetic-deductive method is applied rigorously and hypotheses in question are subjected to careful and robust empirical tests. I do not, by any means, claim, however, that the method suggested here is the only way to study the language faculty. There may be other approaches to the language faculty and there may, and perhaps should, be other types of evidence beside informant intuitions that can be used for or against hypotheses about the language faculty. One might, for example, hypothesize the core property of the language faculty that is different from the model of the Computational System adopted here; one might also have a model of judgment-making that is distinct from what we adopt. The relevant hypotheses and the alternative models in question, however, must be articulated with respect to how informant judgment or other types of evidence, if relevant, is/are related to the hypothesized properties of the language faculty so that we can make testable predictions and aspire to make progress in our endeavor to understand the properties of the language faculty in a research program of language faculty science as an exact science.


Hoji, Hajime. 2009. A foundation of generative grammar as an empirical science. ms, University of Southern California.


Sakuma, Kanae. 1936/1983. *Gendai nihongo no hyoogen to gohoo* (Expressions and usages in modern Japanese). Tokyo: Kurosio Syuppan. (The original 1936 publication was from Koseisy Kooseikaku, Tokyo.)


9. Appendix: Examples with *zibun-zisin* and *kare-zisin*

Section 3.2.2 provides the Schemata in (25) and (26), repeated here.

(25) a. \textit{ok Schema}  
\[\text{NP-ga/wa [NP1-ga zibun-zisin-o/ni V-ru/ta \{to/no ka\} V-ru/ta}\]  
\[\text{NP1 = zibun-zisin}\]

b. \textit{*Schema}  
\[\text{NP1-ga/wa [NP-ga zibun-zisin-o/ni V-ru/ta \{to/no ka\} V-ru/ta}\]  
\[\text{NP1 = zibun-zisin}\]

c. \textit{ok Schema}  
\[\text{NP1-ga/wa [NP-ga \{kare/kanozyo\}-o/ni V-ru/ta \{to/no ka\} V-ru/ta}\]  
\[\text{NP1 = kare/kanozyo}\]

(26) a. \textit{ok Schema}  
\[\text{[[[\alpha-ga zibun-zisin-o/ni V-T to] V-T] NP]-wa ...}\]  
\[\alpha = \text{zibun-zisin}\]

b. \textit{*Schema}  
\[\text{[[[NP-ga zibunzisin-o/ni V-T to] V-T] \alpha]-wa ...}\]  
\[\alpha = \text{zibun-zisin}\]

c. \textit{ok Schema}  
\[\text{[[[NP-ga \{kare/kanozyo\}-o/ni V-T to] V-T] \alpha]-wa ...}\]  
\[\alpha = \text{kare/kanozyo}\]

On the basis of (25) and (26), we can construct the Examples in (60)-(65).

(60) a. \textit{ok Example}  
\[\text{John-wa [Mary-ga zibun-zisin-ni toohyoori-ta to] omoikonde-i-ta}\]  
\'[John thought that Mary had voted for \textit{herself}.']

b. \textit{*Example}  
\[\text{John-wa [Mary-ga zibun-zisin-ni toohyoori-ta to] omoikonde-i-ta}\]  
\'[John thought that Mary had voted for \textit{himself}.']

c. \textit{ok Example}  
\[\text{John-wa [Mary-ga kare-ni toohyoori-ta to] omoikonde-i-ta}\]  
\'[John thought that Mary had voted for \textit{him}.']

(61) a. \textit{ok Example}  
\[\text{John-wa [Mary-ga zibun-zisin-o suisen-i-ta to] bakari omotte-i-ta}\]  
\'[John firmly believed that Mary had recommended \textit{herself}.']
(62) a. *Example
John-wa [Mary-ga zibun-zisin-o suisensi-ta to] bakari omotte-i-ta
'John firmly believed that Mary had recommended himself.'

c. okExample
John-wa [Mary-ga kare-o suisensi-ta to] bakari omotte-i-ta
'John firmly believed that Mary had recommended him.'

(63) a. okExample
Ziro-wa [Hanako-ga zibun-zisin-o hihansi-ta to] it-ta
'Ziro said that Hanako had criticized herself.'

b. *Example
Ziro-wa [Hanako-ga zibun-zisin-o hihansi-ta to] it-ta
'Ziro said that Hanako had criticized himself.'

c. okExample
Ziro-wa [Hanako-ga kare-o hihansi-ta to] it-ta
'John said that Hanako had criticized him.'

(64) a. okExample
'Yoko, who did not even think that John might vote for himself, was very surprised when she learned who John had voted for.'

b. *Example
'Yoko, who did not even think that John might vote for herself, was very surprised when she learned who John had voted for.'

c. okExample
'Yoko, who did not even think that John might vote for her, was very surprised when she learned who John had voted for.'
'John, who firmly believed that Bill would recommend him, was very surprised when he learned who Bill had recommended.'

Similarly, we can construct Examples such as those in (66)-(71) below, on the basis of the Schemata in (27) and (28) with kare-zisin, repeated here.

(27) a. \textit{ok} Schema
\[
\text{NP-ga/wa [NP1-ga kare-zisin-o/ni V-ru/ta \{to/no ka\}] V-ru/ta}
\]
\[
\text{NP1 = kare-zisin}
\]

b. \textit{*} Schema
\[
\text{NP1-ga/wa [NP-ga kare-zisin-o/ni V-ru/ta \{to/no ka\}] V-ru/ta}
\]
\[
\text{NP1 = kare-zisin}
\]

c. \textit{ok} Schema
\[
\text{NP1-ga/wa [NP-ga \{kare/kanozyo\}-o/ni V-ru/ta \{to/no ka\}] V-ru/ta}
\]
\[
\text{NP1 = kare/kanozyo}
\]

(28) a. \textit{ok} Schema
\[
[[[\alpha-ga kare-zisin-o/ni V-T to] V-T] NP]-wa ...
\]
\[
\alpha = \text{kare-zisin}
\]

b. \textit{*} Schema
\[
[[[NP-ga kare-zisin-o/ni V-T to] V-T] \alpha]-wa ...
\]
\[
\alpha = \text{kare-zisin}
\]

c. \textit{ok} Schema
\[
[[[NP-ga kare-o/ni V-T to] V-T] \alpha]-wa ...
\]
\[
\alpha = \text{kare}
\]

(66) a. \textit{ok} Example
\[
\text{Mary-wa [John-ga kare-zisin-ni toohyoosi-ta to] omoikonde-i-ta}
\]
\[
\text{'Mary thought that John had voted for himself.'}
\]

b. \textit{*} Example
\[
\text{John-wa [Mary-ga kare-zisin-ni toohyoosi-ta to] omoikonde-i-ta}
\]
'John thought that Mary had voted for himself.'

c. *Example
John-wa [Mary-ga kare-ni toohyoosi-ta to] omoikonde-i-ta
'John thought that Mary had voted for him.'

(67) a. *Example
Mary-wa [John-ga kare-zisin-o suisensi-ta to] bakari omotte-i-ta
'Mary firmly believed that John had recommended himself.'

b. *Example
John-wa [Mary-ga kare-zisin-o suisensi-ta to] bakari omotte-i-ta
'John firmly believed that Mary had recommended himself.'

c. *Example
John-wa [Mary-ga kare-o suisensi-ta to] bakari omotte-i-ta
'John firmly believed that Mary had recommended him.'

(68) a. *Example
Hanako-wa [Ziro-ga kare-zisin-o hihansi-ta to] it-ta
'Hanako said that Ziro had criticized himself.'

b. *Example
Ziro-wa [Hanako-ga kare-zisin-o hihansi-ta to] it-ta
'Ziro said that Hanako had criticized himself.'

c. *Example
Ziro-wa [Hanako-ga kare-o hihansi-ta to] it-ta
'Ziro said that Hanako had criticized him.'

(69) a. *Example
'Yoko, who did not even think that John might vote for himself, was very surprised when she learned who John had voted for.'

b. *Example
'John, who did not even think that Yoko might vote for himself, was very surprised when he learned who Yoko had voted for.'

c. *Example
'John, who did not even think that Yoko might vote for him, was very surprised when he learned who Yoko had voted for.'

(70) a. *Example
'Yoko, who did not even think that John would recommend himself, was very surprised when she learned who John had recommended.'

b. *Example
   'John, who firmly believed that Yoko would recommend himself, was very surprised when he learned who Yoko had recommended.'

c. *Example
   'John, who firmly believed that Yoko would recommend him, was very surprised when he learned who Yoko had recommended.'

(71) a. *Example
   'Hanako, who learned that Ziro was criticizing himself, was a little surprised.'

b. *Example
   'Ziro, who learned that Hanako was criticizing himself, was a little surprised.'

c. *Example
   'Ziro, who learned that Hanako was criticizing him, was a little surprised.'

A summary of the results of the experiment on the predicted schematic asymmetries in (25) and (26) is given in (72).
A summary of the results of an experiment on (25) and (26), as of March 9, 2010.\(^{80}\)

### Hypothesis: Zibun-zisin is a local anaphor.

<table>
<thead>
<tr>
<th>Schema group 1</th>
<th>the topic construction</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Schema 1 A</td>
<td>51 values</td>
<td>96  (\text{okNP-ga/wa }[^{\alpha}\text{-ga zibun-zisin-o/ni V-T to}]\text{ V-T (under the interpretation that zibun-zisin and }\alpha\text{ are understood to refer to the same individual)})</td>
</tr>
<tr>
<td>Schema 1 B</td>
<td>51 values</td>
<td>64  (\text{*\alpha-ga/wa }[^{\text{NP-ga zibun-zisin-o/ni V-T to}]\text{ V-T (under the interpretation that zibun-zisin and }\alpha\text{ are understood to refer to the same individual)})</td>
</tr>
<tr>
<td>Schema 1 C</td>
<td>51 values</td>
<td>80  (\text{ok\alpha-ga/wa }[^{\text{NP-ga kare-o/ni V-T to}]\text{ V-T (under the interpretation that kare and }\alpha\text{ are understood to refer to the same individual)})</td>
</tr>
</tbody>
</table>

| Schema group 2 | the relative clause construction |  |
|----------------|-------------------------------||--|
| Schema 2 A     | 48 values                      | 99 \(\text{ok[^{[\alpha-ga zibun-zisin-o/ni V-T to}] NP]-wa ... (under the interpretation that zibun-zisin and }\alpha\text{ are understood to refer to the same individual)}\) |
| Schema 2 B     | 48 values                      | 52 \(\text{*[^{[\text{NP-ga zibun-zisin-o/ni V-T to}] V-T} \alpha]-wa ... (under the interpretation that zibun-zisin and }\alpha\text{ are understood to refer to the same individual)}\) |
| Schema 2 C     | 38 values                      | 67 \(\text{ok[^{[\text{NP-ga kare-o/ni V-T to}] V-T} \alpha]-wa ... (under the interpretation that kare and }\alpha\text{ are understood to refer to the same individual)}\) |

17 participants  
538 answers

As in the case of the *Schema-based prediction in (15a) about otagai, the *Schema-based prediction in (15b) about zibun-zisin is clearly disconfirmed.

A summary of the results of the experiment on the predicted schematic asymmetries in (27) and (28) is given in (72).

---

\(^{80}\) There are three Examples for each Schema in (72); see (25)-(65). "51 values" in (73) means that every informant has reported his/her judgment on at least one Example on each Schema of the topic construction.
(73) A summary of the results of an experiment on (27) and (28), as of March 27, 2010:  

**Hypothesis:** *Kare-zisin* is a local anaphor.

<table>
<thead>
<tr>
<th>Schema group 1</th>
<th>in the topic construction</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Schema 1 A</strong></td>
<td>48 values 92</td>
<td>$^{\text{ok}}$NP-ga/wa $^{\text{[a-ga}}$ *kare-zisin-o/ni V-T to] V-T (under the interpretation that <em>kare-zisin</em> and $^\alpha$ are understood to refer to the same individual)</td>
</tr>
<tr>
<td><strong>Schema 1 B</strong></td>
<td>48 values 58</td>
<td>$^{*}$NP-ga/wa $^{\text{[NP-ga}}$ *kare-zisin-o/ni V-T to] V-T (under the interpretation that <em>kare-zisin</em> and $^\alpha$ are understood to refer to the same individual)</td>
</tr>
<tr>
<td><strong>Schema 1 C</strong></td>
<td>48 values 89</td>
<td>$^{\text{ok}}$NP-ga/wa $^{\text{[NP-ga}}$ *kare-o/ni V-T to] V-T (under the interpretation that <em>kare</em> and $^\alpha$ are understood to refer to the same individual)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Schema group 2</th>
<th>in the relative clause construction</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Schema 2 A</strong></td>
<td>48 values 93</td>
<td>$^{\text{ok}}$[[NP-ga]} *kare-zisin-o/ni V-T to] V-T] NP]-wa ... (under the interpretation that <em>kare-zisin</em> and $^\alpha$ are understood to refer to the same individual)</td>
</tr>
<tr>
<td><strong>Schema 2 B</strong></td>
<td>48 values 60</td>
<td>$^{*}$[[NP-ga]} *kare-zisin-o/ni V-T to] V-T] $^\alpha$]-wa ... (under the interpretation that <em>kare-zisin</em> and $^\alpha$ are understood to refer to the same individual)</td>
</tr>
<tr>
<td><strong>Schema 2 C</strong></td>
<td>48 values 73</td>
<td>$^{\text{ok}}$[[NP-ga]} *kare-o/ni V-T to] V-T] $^\alpha$]-wa ... (under the interpretation that <em>kare</em> and $^\alpha$ are understood to refer to the same individual)</td>
</tr>
</tbody>
</table>

16 participants  
477 answers

The *Schema-based prediction in (15c) about *kare-zisin* is thus also clearly disconfirmed. We had conducted experiments a few years ago under a general experimental design that was not as systematic as what is reported above, checking informant judgments on examples that conformed to (27b), given then as "control." In one of those experiments, such examples conforming to (27b) were judged to be unacceptable consistently only by a few informants out of over 60 informants. In another such experiment, all of the 11 informants accepted examples conforming to (27b).