Hypothesis testing in generative grammar: Evaluation of predicted schematic asymmetries[*]

Hajime Hoji
University of Southern California

This paper explores how the *hypothetico-deductive* method can be applied to research concerned with the properties of the language faculty. The paper first discusses how we can try to identify informant judgments that are likely a reflection of properties of the Computational System (or properties of the language faculty that are directly related to the Computational System), proposes a method of hypothesis testing in line with the *hypothetico-deductive* method, and provides an illustration by examining the predictions made under the lexical hypothesis that *otagai* in Japanese is a local anaphor.

Areas of interest: syntax, methodology, local anaphors, Japanese

1. Introduction

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

> 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)

Feynman continues by adding the following "obvious remarks":[1]

[*] 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 about the content in February 2009, 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, 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; so have comments by Chris Kennedy and Hiroki Narita on a related work. Discussions with Ayumi Ueyama and Yuki Takubo over the years have helped me form the views presented here. I would also like to thank JJ Nakayama for his comments. None of the aforementioned is responsible for the remaining errors and shortcomings of the paper. Earlier versions of sections 2-3 of the present paper have appeared in Hoji (2010).

[1] 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 5.3 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).

> 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)

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.

(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 to be 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. The research that underlies this paper, however, pursues the thesis that we can make point-value predictions in language faculty science (and we in fact should, given the conception of the language faculty and the research heuristics adopted here) although the page limit does not allow me to provide a *full* illustration of how that can be done.

Section 2 addresses methodological issues and makes a proposal for testing our hypotheses about properties of the language faculty. Section 3 provides a brief illustration of the proposal. Section 4 offers a summary of the proposed methodology, making reference to two research heuristics, and some implications are discussed in Section 5. The language dealt with is Japanese. The late Yuki Kuroda's research has been guided by the belief that hypotheses about the language faculty can be tested by careful experiments dealing with (a) particular language(s).[2] The work presented here tries to follow in his footsteps.[3]

2. Proposal

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.[4] 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*

---

[2] Kuroda (1999: section 5 普遍文法は個別文法を実験室として仮説検証法によって追求される [The discovery of the properties of UG is sought in the laboratory of a particular grammar by the hypothesis-testing method, HH]), contains some relevant remarks.

[3] The methodological proposal in this paper is based on Hoji (2009). A more complete discussion and further illustration will be provided in separate works.

[4] 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.

Linguistic Analysis in English, May 9-12, 1958, published in 1962 by the University of Texas.[5]

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 understood in generative research that the Computational System is an algorithm whose input is a set of items taken from the mental Lexicon of speakers of a language and whose output is 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:

Numeration μ => [CS] => LF(μ)
⇓
PF(μ)

Numeration μ: a set of items taken from the mental Lexicon
LF(μ): an LF representation based on μ
PF(μ): a PF representation based on μ

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. The 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 therefore be understood as demonstrating the existence of the Computational System by discovering its properties.[6]

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 commit 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

---

[5] 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?
Chomsky: Because I am interested in explaining intuition. If you cannot accept this as the purpose of linguistic study, I am lost. I would like to get a theory which will predict intuitions.

Obviously, informant judgments are not the only source of evidence for or against hypotheses about the Computational System. If one seeks evidence elsewhere, one must articulate how such 'evidence' is related to the hypothesized properties of the language faculty so as to ensure, and hopefully maximize, testability of the hypotheses. I take that to be a minimal methodological requirement for using evidence other than informant intuitions in empirical research concerned with the properties of the Computational System.

[6] 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).

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:

γ(a, b)

α ≈≈> [μ => CS => LF(μ)] ≈≈> β
⇓
PF(μ)

a. γ(a, b): an intuition that two linguistic expressions *a* and *b* are related in a particular manner[7]
b. α: the presented sentence
c. β: the informant judgment on the acceptability of α under γ(a, b)

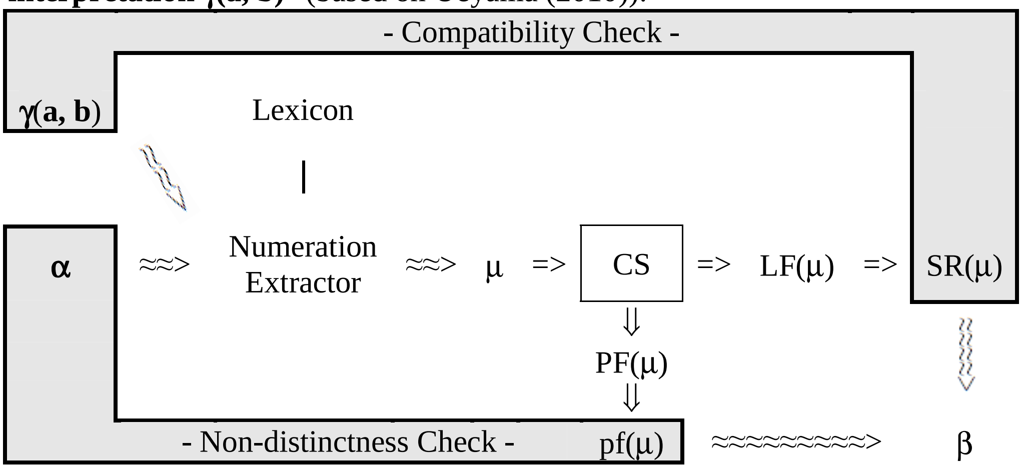
The boxed part in (3) is the Computational System; see (2). The informant is presented sentence α and asked whether it is acceptable, or how acceptable it is, under a particular interpretation γ(a, b) involving two linguistic expressions *a* and *b*. As noted above, insofar as informant judgments are 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 μ and compares (i) the two output representations based on μ with (ii) the 'sound' (i.e., the presented sentence α) and the relevant 'meaning' under discussion (i.e., the interpretation γ(a, b)).

The following model of judgment-making by informants presents itself.[8]

[7] 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 γ, γ', and γ'', they can be expressed as γ(John, himself) in *John praised himself*, γ'(everyone, his) in *everyone praised his family*, and γ''(everyone, someone) in *everyone praised someone*. See footnote 9 below.

[8] 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, as Ueyama points out, 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). Hoji (2009: Appendix) compares (4) with the model of judgment-making suggested in Schütze (1996: 175).

(4) The Model of Judgment-Making by the Informant on **the acceptability of sentence α with interpretation γ(a, b)**[9] (based on Ueyama (2010)):



a. α: the presented sentence
b. μ: numeration
c. γ(a, b): the interpretation intended to be *included* in the 'meaning' of α involving expressions *a* and *b*
d. LF(μ): the LF representation that obtains on the basis of μ
e. SR(μ): the information that obtains on the basis of LF(μ)
f. PF(μ): the PF representation that obtains on the basis of μ
g. pf(μ): the surface phonetic string that obtains on the basis of PF(μ)[10]
h. β: the informant judgment on the acceptability of α under γ(a, b)

The "==>" 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 "≈≈>", on the other hand, is not an input/output relation, as roughly indicated in (5).[11]

(5)a. Presented Sentence α ≈≈> Numeration Extractor: ... contributes to ...
b. γ(a, b) ≈≈> Numeration Extractor: ... contributes to ...
c. Numeration Extractor ≈≈> numeration μ: ... forms ... [12]
d. γ(a, b) – Compatibility Check – SR(μ) ≈≈> Judgment β: ... serves as a basis for ... [13]
e. α - Non-distinctness Check - pf(μ) ≈≈> Judgment β: ... serves as a basis for ...[14]

---

9 It is argued in some depth in Hoji (2009: chapter 5) that it would be qualitatively more difficult to maximize our chances of learning from errors if we dealt with simple (un)acceptability *without* considering γ(a, b). Due to space limit, I cannot provide the relevant arguments in this paper; but see section 4.2 below.

10 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.

11 As discussed in some depth in Hoji (2009), the model of judgment-making in (4) can be regarded as a consequence of adopting the theses, shared by most practitioners of generative grammar, that the Computational System in (2) is at the center of the language faculty and that informant judgments are a primary source of evidence for or against our hypotheses pertaining to properties of the Computational System.

12 It is assumed that "Numeration Extractor" makes reference to "Lexicon."

13 Compatibility between SR(μ) and γ(a, b) or the lack thereof serves as a basis for β.

14 Non-distinctness between pf(μ) and α or the lack thereof also serves as a basis for β.

2.3. Informant judgments and the fundamental asymmetry

Crucial for making testable predictions is a claim—which is called in Hoji (2009) a *bridging statement*—that γ(a, b) (see (4b) and footnote 7) 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 β 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 μ that would result in pf(μ) non-distinct from α and SR(μ) compatible with the interpretation γ(a, b), that may not necessarily result in the informant reporting that α is (fully) acceptable under γ(a, b). On the other hand, if the informant fails to come up with such a numeration μ, the informant's judgment on α under γ(a, b) *should* necessarily be "complete unacceptability". For, in that case, the informant fails to "arrive at" SR(μ) compatible with the interpretation γ(a, b) presumably because the hypothesized structural condition necessary for γ(a, b) is never met in *any* LF(μ) no matter what possible μ might be tried. This is the source of the fundamental asymmetry between a **Schema*-based prediction and an ^{ok}*Schema*-based prediction (to be discussed 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, and hence 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, basic units of facts, and *confirmed schematic asymmetries*

The minimal empirical prerequisite for effective pursuit of the discovery of the properties of the language faculty is being able to identify informant intuitions that are a likely reflection of properties of the Computational System hypothesized to be at the center of the language faculty. Without being able to identify what is a likely reflection of properties of the Computational System, 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".

It is proposed in Hoji (2009) 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. In Hoji (2009), the former type of Schema is called a **Schema* (which can be read as "star schema") and sentences conforming to it are called **Examples* (which can be read as "star examples") and the latter type of Schema is called an ^{ok}*Schema* and sentences conforming to it are called ^{ok}*Example*s.

A **Schema*-based prediction is as in (6), and one of the possible formulations of an ^{ok}*Schema*-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

[15] *Predictions* here cover both predictions and retrodictions.

interpretation γ(*a*, *b*).

(7) An ^{ok}*Schema*-based prediction—version 1:[16]
Informants judge ^{ok}*Examples* conforming to an ^{ok}*Schema* to be acceptable (to varying degrees) under interpretation γ(*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 ^{ok}*Schema*-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 ^{ok}*Schema*-based prediction. Instead of (7), one can adopt (8), for example, which is less stringent than (7) because the existence of *just one* ^{ok}*Example* that is judged to be acceptable would confirm (8).

(8) An ^{ok}*Schema*-based prediction—version 2:
Informants judge *some* ^{ok}*Example* conforming to an ^{ok}*Schema* to be acceptable (to varying degrees) under interpretation γ(*a*, *b*).

If we adopt the formulation of an ^{ok}*Schema*-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: ^{ok}*Schema*-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*, provided that there is at least some freedom as to how a given Schema gets instantiated by a particular Example, with regard to the choice of particular lexical items or the addition of an optional phrase (which is reasonable given that we are concerned with properties of the Computational System, and also in light of the "Maximize testability" heuristic, to be addressed below).

For the reasons noted above, we should expect complete unacceptability if there is no numeration μ corresponding to α that would result in (i) LF(μ) (hence SR(μ)) compatible with γ(a, b) and (ii) PF(μ) (hence pf(μ)) non-distinct from α. The content of a **Schema*-based prediction is that there is no such numeration. The informant judgment that α is *not* completely unacceptable under γ(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 μ corresponding to α that would result in LF(μ) (hence SR(μ)) compatible with γ(a, b) and PF(μ) (hence pf(μ)) non-distinct from α. α's not-fully-acceptable status under γ(a, b) must be due to extra-grammatical factor(s).[17] 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).[18]

---

[16] We will consider below two other possible formulations of an ^{ok}*Schema*-based prediction.

[17] What is disregarded here 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. We should try to minimize such a possibility by improving the design of the experiment and also conducting preliminary experiments. The space limit, however, does not allow me to illustrate how that can be, and has been, done.

[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 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

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 basic unit of facts 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 minimally differ from the**Examples* in terms of the hypothesized formal property are acceptable under $\gamma(a, b)$ (at least to some extent) and (ii) ^{ok}*Examples* that are identical to the **Examples* but do not involve interpretation $\gamma(a, b)$ 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.[19]

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[20] and that of the corresponding ^{ok}*Schemata* is higher than 0.[21] 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 should impose on our ^{ok}*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 ^{ok}*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

---

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 likely to be independent of parsing difficulty.

[20] 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" (i.e., the average score/value) of the **Schema* among the entire informants, on the scale of "0" (for complete unacceptability) to "100" (for full acceptability). 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; see footnote 49 below.

[21] The "representative value" of a Schema is computed based on the informant judgments on the Examples that conform to the Schema; see Ueyama (2010) for more details. In what follows, the "average score/value" is sometimes used instead of the "representative value."

matter to determine what the "representative value" of the ^{ok}*Schemata* should be in order for a *confirmed schematic asymmetry* to obtain, the researchers themselves perhaps should aspire to the "standard" suggested in (9) below, leaving aside its actual feasibility in every single experiment.[22]

(9) An ^{ok}*Schema*-based prediction—version 3:
Informants judge *every* ^{ok}*Example* (in an experiment) conforming to an ^{ok}*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 done by Tycho Brahe and other such observations in physical sciences; see Feynman's 1965/94 remarks below.[23]

> ... 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)

Given that "[i]t is only through such hard work that we can find out anything", it is clear that we should

[22] Hoji (2009) provides a great deal more discussion on the relevant issues, making reference to concepts such as informant *resourcefulness*, *single-informant* and *multiple-informant experiments*. One of the points made there is 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*.

[23] 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 ^{ok}*Schemata* are based on various hypotheses about properties of the Computational System, those about how a certain type of informant intuition arises based on certain properties at LF—they are called in Hoji (2009) *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 perhaps closer to observations that have been replicated with the aid of various observation devices. It is independent theories (e.g., a theory of optics in the case of telescopes) that determine how such devices work and how we interpret what is "observed through such devices."

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, 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 ^{ok}*Schema*-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 ^{ok}*Schema*-based predictions have been confirmed. Does this mean that we are justified to conclude that we now have a *confirmed schematic asymmetry*? Strictly speaking, 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. While 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, it is still possible that the researcher has failed to, but someone else can, come up with **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* (along with ^{ok}*Examples*) 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, for example, 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. 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/value 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; and 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 ^{ok}*Schema*-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.[24, 25]

[24] 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. As discussed in section 4.2 below, 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

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. 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.[26]

(13) NP1 Verb [that NP2 Verb NP3]

That is to say, if one puts forth or accepts a hypothesis that expression α 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 α are unacceptable if α 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. ^{ok}*Schema*
 NP V himself

---

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 in such cases on the basis of hypotheses that might have been “accepted” despite the absence of any compelling empirical justification.

[25] 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.

[26] The use of “NP” in place of “DP” is inconsequential in this paper.

NP=himself

b. **Schema*
NP1 V that NP2 V himself
NP1=himself

c. ^{ok}*Schema*
NP1 V that NP2 V him
NP1=him

As suggested above, what is predicted is a *schematic asymmetry*; more specifically, the prediction is that there are no Examples conforming to (14b) that are judged not completely unacceptable while there are Examples conforming to (14a) and (14c) that are judged (more or less) acceptable under the interpretations indicated in (14a) and (14c). We are not going to address in this paper how robust the informant judgments are on this predicted *schematic asymmetry*; we will 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 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).[27]

(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
'NP1 Verb that NP2 Verb NP3'

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. Due to space limitation, we will only discuss (15a).

---

[27] 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).

3.2.2. **Schema*-based predictions and ^{ok}*Schema*-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.

(17)a. ^{ok}*Schema*
NP-ga/wa [$_{CP}$ NP1-ga otagai-o/ni V-ru/ta {to/no ka}] V-ru/ta
under the reciprocal reading of *otagai* with NP1 as its "antecedent"
b. **Schema*
NP1-ga/wa [$_{CP}$ NP-ga otagai-o/ni V-ru/ta {to/no ka}] V-ru/ta
under the reciprocal reading of *otagai* with NP1 as its "antecedent"
c. ^{ok}*Schema*
NP1-ga/wa [$_{CP}$ NP-ga karera-o/ni V-ru/ta {to/no ka}] V-ru/ta
under the coreference between *karera* and NP1

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 ^{ok}*Schema*. Because *karera* 'them' is not marked [+A], (17c) is also an ^{ok}*Schema*.

In each Schema in (18), 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 ^{ok}*Schemata*.

(18)a. ^{ok}*Schema*
[$_{NP}$ [$_{TP}$ *ec* otagai-o/ni V-ru/ta] *NP1*]
under the reciprocal reading of *otagai* with NP1 as its "antecedent"
b. **Schema*
[$_{NP}$ [$_{TP}$ *ec* [$_{CP}$ NP-ga otagai-o/ni V-ru/ta {to/no ka}] V-ru/ta] *NP1*]
under the reciprocal reading of *otagai* with NP1 as its "antecedent"
c. ^{ok}*Schema*
[$_{NP}$ [$_{TP}$ *ec* [$_{CP}$ NP-ga karera-o/ni V-ru/ta {to/no ka}] V-ru/ta] *NP1*]
under the coreference between *karera* and NP1

On the basis of the Schemata in (17), we can construct the Examples in (19) and (20).[28]

(19)a. ^{ok}*Example*
Mary-wa [*John to Bill*-ga *otagai*-ni toohyoosi-ta to] omoikonde-i-ta
'Mary thought that *John and Bill* had voted for *each other*.'
b. **Example*
John to Bill-wa [Mary-ga *otagai*-ni toohyoosi-ta to] omoikonde-i-ta
'*John and Bill* thought that Mary had voted for *each other*.'
c. ^{ok}*Example*
John to Bill-wa [Mary-ga *karera*-ni toohyoosi-ta to] omoikonde-i-ta

[28] Hoji (2006) contains further examples and discussion.

‘*John and Bill* thought that Mary had voted for *them*.’

(20)a. ^{ok}*Example*

Sensei-wa [*John to Bill*-ga naze *otagai*-o suisensi-ta no ka] mattaku wakara-nakat-ta
‘The teacher had no idea why *John and Bill* had recommended *each other*.’

b. **Example*

John to Bill-wa [sensei-ga naze *otagai*-o suisensi-ta no ka] mattaku wakara-nakat-ta
‘*John and Bill* had no idea why the teacher had recommended *each other*.’

c. ^{ok}*Example*

John to Bill-wa [sensei-ga naze *karera*-o suisensi-ta no ka] mattaku 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. ^{ok}*Example*

[[*ec* sensyuu-no senkyo-de *otagai*-ni toohyoosi-ta] *John to Bill*]-wa Susan-ga dare-ni toohyoosi-ta ka sit-te odoroi-ta.
‘*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*

[[*ec* [[Susan-ga sensyuu-no senkyo-de *otagai*-ni toohyoosi-ta] to] omoikonde-i-ta] *John to Bill*]-wa Susan-ga dare-ni toohyoosi-ta ka sit-te odoroi-ta.
‘*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. ^{ok}*Example*

[[*ec* [[Susan-ga sensyuu-no senkyo-de *karera*-ni toohyoosi-ta] to] omoikonde-i-ta] *John to Bill*]-wa Susan-ga dare-ni toohyoosi-ta ka sit-te odoroi-ta.
‘*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. ^{ok}*Example*

[[*ec* kondo-no yakusyoku-ni *otagai*-o suisensi-ta] *John to Bill*]-wa iroirona hito-ni meeru-o okut-te riyuu-o setumeisi-te-i-ru rasii.
‘I hear that *John and Bill*, who have recommended *each other* for the new post, are emailing various people to explain why.’

b. **Example*

[[*ec* [Mike-ga kondo-no yakusyoku-ni naze *otagai*-o suisensi-ta no ka] siritagat-te-i-ta] *John to Bill*]-wa iroirona hito-ni meeru-o okut-te riyuu-o sirabe-te-i-ru rasii.
‘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. ^{ok}*Example*

[[*ec* [Mike-ga kondo-no yakusyoku-ni naze *karera*-o suisensi-ta no ka] siritagat-te-i-ta] *John to Bill*]-wa iroirona hito-ni meeru-o okut-te riyuu-o sirabe-te-i-ru rasii.
‘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 ^{ok}*Schema*-based prediction:
The ^{ok}*Examples* conforming to the ^{ok}*Schemata* in (17a), (18a), (17c) and (18c) are *not* completely unacceptable, including the (a) and (c) examples in (19)-(22).

3.3. Experiments

One can test **Schema*-based predictions and corresponding ^{ok}*Schema*-based predictions by checking informant judgments on **Examples* and the corresponding ^{ok}*Examples*, to see if we obtain a *confirmed schematic asymmetry*. Here, I would like to briefly introduce the general design of experiments that we have been conducting.[29]

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 (25), for example, once translated into English.

(25)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 (26). And what the informant has indicated is converted to numerical values as in (27), 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.

(26) Bad ⇔ Good
o o o o o
(27) 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.

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

[29] I should like to acknowledge that our on-line experiments are based on the program that has been created by Ayumi Ueyama.

(28) A summary of the results of the experiment as of June 14, 2010:

Schema group 1	*Otagai* is in the embedded object position.			
	Schema 1 A (=(17a))	68 values	96	^{ok}*Examples* in (19a) and (20a)
	Schema 1 B (=(17b))	67 values	61	**Examples* in (19b) and (20b)
	Schema 1 C (=(17c))	68 values	85	^{ok}*Examples* in (19c) and (20c)
Schema group 2	*Otagai* is in the embedded object position. The intended antecedent is the relative head.			
	Schema 2 A (=(18a))	68 values	94	^{ok}*Examples* in (21a) and (22a)
	Schema 2 B (=(18b))	68 values	53	**Examples* in (21b) and (22b)
	Schema 2 C (=(18c))	68 values	70	^{ok}*Examples* in (21c) and (22c)
34 informants, 865 answers				

“Schema group 1” is for (17) and “Schema group 2” is for (18). “Schema 1 A” covers the ^{ok}*Examples* in (19a) and (20a), “Schema 1 B” the **Examples* in (19b) and (20b), and “Schema 1 C” the ^{ok}*Examples* in (19c) and (20c). Likewise, “Schema 2 A” covers the ^{ok}*Examples* in (21a) and (22a), “Schema 1 B” the **Examples* in (21b) and (22b), and “Schema 1 C” the ^{ok}*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 (28) 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 (28)). The average scores on (19b) and (20b) are 50 and 72, respectively, and “61” for Schema 1B in (28) is the average of those two scores.

The average scores/values 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 (28) thus clearly disconfirm the **Schema*-based predictions based on the lexical hypothesis in (15a), repeated here.[31,32]

---

[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 (28)); but the former also judge ^{ok}*Examples* less acceptable than the latter (e.g., 63 by the former and 88 by the latter on Schemata C in (28)). This is expected from the considerations of informants’ *resourcefulness* as discussed in Hoji (2009).

[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. The mean “scores” on these two types of Examples, as of June 14, 2010, are 60 (by 20 informants) and 93 (by 19 informants), contrary to the predicted 0. Likewise, the **Schema*-based predictions in accordance with the 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.

As H.-D. Ahn (p.c., 12/12/2009) suggests, one might pursue the possibility that the hypotheses in (15) are valid but that *otagai*, *zibun-zisin* and *kare-zisin* always occur in a structural position in which they have a covert antecedent in its local

(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). Several such attempts are in fact discussed in Hoji (2009) and it is concluded there that such attempts either end up being *content-reducing* (i.e., *degenerating*) in the terms of Lakatos (1970/1978) — resulting only in the elimination of the **Schema*-based prediction without introducing a new **Schema*-based prediction — or result in the disconfirmation of the new **Schema*-based prediction. 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), according to which 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, 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.

---

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* (i.e., *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 (28) 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 (28) 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 is perhaps worth noting the following: Of the 34 informants in the experiment whose results are summarized in (28) 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 (28) are 0 and 3. (The other 32 informants accept **Examples* of the **Schemata* in (28) 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 wind made each other feel very happy" are 87 and 75, respectively. Hence, one cannot maintain that (15a) is valid for that informant. The other informant has not participated in the experiments that contain such examples.

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?[34] 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 ^{ok}*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.[35] 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.[36]

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

(29) The two aspects of the "Maximize testability" heuristic:
  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 36.

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 (29b), 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 (29b).[37]

I should like to note that (29b) is also applicable when we modify our hypotheses in response to the failure of our predictions. That is to say, modification of our hypotheses should proceed in line with

---

[34] I will return to the use of statistics in section 5.1.

[35] But see section 2.4 for the importance of ^{ok}*Schema*-based predictions.

[36] As Yasuo Deguchi (p.c. May 2010) points out, 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, for example, by counting various sub-cases of a more general **Schema*-based prediction as "distinct" **Schema*-based predictions. 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.

[37] 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).

(29b). 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 errors[38]

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".[39]

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".[40] We can try to maximize our chances of learning from errors by minimizing the number of hypotheses whose validity is to be tested in a given experiment, and more importantly, also 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 testing the predictions made under (15a), (11) and the hypotheses that ensure (30).[41]

(15)a. *Otagai* is marked [+A].
(11) A [+A] category must have an antecedent in its local domain.
(30) 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'

---

[38] 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 will only make this qualification and continue to use what is given as the heading of this subsection. Various related issues will have to be addressed in separate works.

[39] 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.

[40] 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.

[41] 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 empirical invalidation.

In accordance with Duhem's thesis noted above, the disconfirmation of the **Schema*-based prediction 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. Similarly, the validity of (30) is widely accepted among the researchers who address how (11) applies to Japanese. The validity of both (11) and (30) thus allows us to 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 (30), 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), 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 (30).

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. 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.[42]

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.

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

[42] See Poincaré (1952: chap. 9, 151-152) for relevant discussion. The general point 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 anaphors" but also in regard to variable binding, quantifier scope, "floating quantifiers," and other "empirical domains," which, as suggested above, seems to be related to the failure to understand the significance of making **Schema*-based predictions. The illustration of this, however, will have to be made on a separate occasion due to space considerations.

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.[43] 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; but see note 38.[44]

5. Some implications

5.1. 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.[45] 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

---

[43] Recall that **Schema*-based predictions can, but *^{ok}Schema*-based predictions cannot, be disconfirmed.

[44] In the foregoing discussion, we 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 7. I noted in footnote 9 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). I would like to briefly state why.

It is not possible to attribute the complete unacceptability of **Example* α to "parsing" difficulty or the unnaturalness of the interpretation of the entire α, as long as its *two* corresponding *^{ok}Examples* are judged (more or less) acceptable. The complete unacceptability of α can thus be reasonably be attributed to the hypothesized condition(s) for γ(a, b) not being satisfied in any LF that could possibly correspond to α. (Even marginal) acceptability of **Example* α, on the other hand, can be taken as evidence that at least one of the hypotheses that give rise to the relevant **Schema*-based prediction is not valid, *insofar as* we can ensure, on the basis of the relevant informants' judgments in a preliminary experiment, that the informants clearly understand what is intended by γ(a, b) in question and the instructions given in the experiment. (It is, in principle, possible to have such a preliminary experiment as long as γ(a, b) is based not only on a universal condition but also on a language-particular condition because an experiment on the language-particular condition can serve as a preliminary experiment for the one that involves the universal condition as well as the language-particular condition.

Without γ(a, b), we do not have two *^{ok}Examples* corresponding to the **Example*. Unlike the case *with* γ(a, b) considered above, we cannot therefore, in principle, have a reasonable ground for excluding the complete unacceptability of a **Example* to "parsing" difficulty or the unnaturalness of the interpretation of the entire α. Hence the complete unacceptability in question cannot be reasonably regarded as being due to the hypothesized grammatical reason. Likewise, it is not clear how we can take (even marginal) acceptability of a **Example* as evidence that at least one of the hypotheses that give rise to the **Schema*-based prediction is not valid because it is not clear how we can ensure, on the basis of the relevant informants' judgments in a preliminary experiment, that the informants can distinguish "grammatical well-formedness" and intelligibility, so to speak.

[45] We can make use of a "standard" statistical method and present the results of our experiments, graphically or otherwise, by utilizing SPSS (Statistical Package for the Social Sciences), for example. It is true that the graphic representation of the results makes it visually transparent between cases of the (blatant) failure to obtain a *confirmed schematic asymmetry* and those more successful cases, but the application of the "standard" statistical method does not alter the verdict that a given **Schema*-based prediction has been disconfirmed where the average score/value of the **Schema* is substantially higher than 0.

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 ^{ok}*Schemata*, on the 0-100 scale, does not count as a *confirmed schematic asymmetry* according to the proposal suggested above. We adopt the thesis 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'. And this separates the present approach, in a crucial way, from most other "experimental approaches" in the field, including Cowart (1997) and other works in the Magnitude Estimation approach; cf. Sprouse (2007: 124) and Young (2008: 209-10) for some relevant discussion.[46]

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' 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 note 46. 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 expect them to be supported by experimental results, 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.

---

[46] Morrison and Henkel (1970/2007) contains works that critically address the use of significance tests in social sciences, 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 $\bar{x}$, 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. 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.

5.2. On cross-linguistic research

Hypotheses about the Computational System are universal and are necessarily quite abstract. Actual experiments to test the validity of those hypotheses, however, must be concrete because they must deal with specific predictions in *a particular language*. This is the source of a fundamental challenge in trying to ensure the testability of hypotheses about *the Computational System*. For example, the empirical consequences of hypotheses about the Computational System that have been backed up by a *confirmed schematic asymmetry* in one language should in principle be testable with respect to any other language, but lexical differences among languages can 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*.

5.3. 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 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.[47] 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 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*.[48]

---

[47] For example, the value of the anomalous precession of Mercury's perihelion 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.

[48] 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*.

6. Concluding remarks

The starting points of this paper were the assumptions (i) that the main goal of our research in generative grammar is to discover the properties of the Computational System, hypothesized to be at the center of the language faculty, and (ii) that a major source of evidence for or against our hypotheses is informant judgments; see section 2.1. Despite a wide acceptance of (ii), 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."[49]

In this paper I have proposed a means to identify informant judgments that are likely a reflection of properties of the Computational System, arguing that our research endeavor must be built on *confirmed schematic asymmetries*, which I have suggested 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 **Examples* are consistently "completely unacceptable" and their judgments on the corresponding ^{ok}*Examples* 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.[50]

One might wonder if this is an unrealistically high standard because we cannot fully control various non-grammatical factors. In response to such a possible objection, I should like to mention that it is in fact possible to obtain a *confirmed schematic asymmetry* in accordance with what has been suggested above. In experiments we have conducted on bound variable anaphora, split antecedence, and local disjointness (i.e., so-called Binding Principle B) effects in Japanese, for example, the mean score/value of the **Schemata* are around, or lower than, 5 while those of many of the corresponding ^{ok}*Schemata* are over 90, on the 0-100 scale.

An illustration of the proposed method should in fact been a great deal more effective if we discussed lexical and structural hypotheses that seem to be supported by *confirmed schematic asymmetries* in Japanese, along with the relevant experimental results. Without such an illustration and demonstration, the empirical discussion in this paper could be taken as further support for the thesis that it is not possible to obtain informant judgments of a categorical nature; but that is quite contrary to, and

---

[49] Schütze (1996) contains extensive literature review; cf. also Devitt (2006), and Fitzgerald (2010) for recent discussion. 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 imaginable 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.

[50] 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 **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 a lexical property).

is in fact the opposite of, my contention. Here I should therefore like to briefly report on the result of an experiment that indicates that it is possible to obtain a *confirmed schematic asymmetry* by adhering to the "high standard" suggested above.[51]

Japanese has three non-interrogative demonstrative prefixes *ko*- 'this', *so*- 'that', *a*- 'that'.[52] While both *so*-NPs and *a*-NPs can be anaphorically related with another NP, it has 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. The anaphoric relation in question has been considered as an instance of bound variable anaphora. The generalization is recorded in (31) and is illustrated in (32).

(31) *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'.

(32)a. *Kanarino kazu-no seizika*-ga {*so/*a*}*-itu/*-no hisyo-o hihansita.
considerable number-GEN 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 zassikisya]-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 (31) in theoretical terms, the examples with a *so-NP* in (32) conform to the *okSchema* in (33a) and those with an *a-NP* in (32) to the **Schema* in (33b).[53]

(33) (Where *NP1* is not singular-denoting)

a. *okSchema*:
NP1-ga ... *so-NP* ... V
BVA(NP, so-NP)

b. **Schema*:
NP1-ga ... *a-NP* ... V
BVA(NP, a-NP)

c. *okSchema*:
NP1-ga ... a-NP ... V
(With *a*-NP "referring to" a particular individual/object.)

As of May 27, 2010, the average scores/values of the *Schemata* in (33a), (33b) and (33c) are 75, 5, 94, respectively, with 37 informants (15 "linguistically informed" and 22 "linguistically naïve"). The **Schema*-based prediction involving (33b) thus seems to have survived a disconfirmation attempt, in sharp contrast with the **Schema*-based prediction based on (15a).[54]

---

51 A more in-depth illustration of the proposed methodology will have to be made in separate works.

52 The standard literature on demonstratives in Japanese includes works by Matsushita, Sakuma, Mikami, Kuno and Kuroda.

53 There must be a language-specific lexical hypothesis that yields the effects of (31); Ueyama (1998) puts forth a formal proposal that has precisely those effects.

54 The result of a similar experiment reported in Aoshima et al. (2009: 3.2) is substantially different from what is reported here; once converted to the 0-100 scale, the mean scores on their **Schemata* are over 20 and close to 25, with 28 informants (all undergraduate students). The discrepancy between the results of our experiment and theirs raises interesting questions in

To summarize, 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, given that we want our hypotheses to be testable in accordance with the *hypothetico-deductive* method, but it is also a consequence of adopting the model of judgment-making in (4) and incorporating a *bridging statement* that states a necessary condition for a particular interpretation involving two linguistic expressions *a* and *b*, γ(a, b); see section 2.2, including footnote 7. The present work maintains, in the terms of Nakaya's characterization of science (see section 4.1 above), that basic units of facts in language faculty science are *confirmed schematic asymmetries*, which are *reproducible phenomena* that are "measurable", so to speak. 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 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. Most importantly, it makes us hopeful that we might be able to make generative grammar an empirical science, or to put it more generally, to make *language faculty science* possible, where the general scientific method in (1) can be rigorously applied.[55]

References

Aoshima, Sachiko, Yoshida, Masaya, and Phillips, Collin. 2009. Incremental processing of coreference and binding in Japanese. Syntax 12.2: 93-134.

Ariew, Roger. 1984. The Duhem thesis. The British Journal for the Philosophy of Science 35: 313-325.

---

relation to the precision of our measuring devices, so to speak, which include various aspects of the general experimental design including how "instructions" are given to the informants. The relevant issues will be addressed in a separate work.

[55] 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 *hypothetico-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 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.

Barnard, Chris, Gilert, Francis, and McGergor, Peter. 2007. Asking questions in biology: a guide to hypothesis-testing, experimental design and presentation in practical work and research projects. Third edition. Essex, England: Pearson Education Limited.

Boeckx, Cedric. 2006. Linguistic minimalism: Origins, concepts, methods, and aims. New York: Oxford University Press.

Chomsky, Noam. 1965. Aspects of the theory of syntax. Cambridge, MA: MIT Press.

Chomsky, Noam. 1981. Lectures on government and binding. Dordrecht: Foris.

Chomsky, Noam. 1986. Knowledge of language. New York: Praeger.

Chomsky, Noam. 1993. A Minimalist Program for linguistic theory. In: Hale, Kenneth, and Samuel Jay Keyser (eds.), The view from Building 20: Essays in linguistics in honor of Sylvain Bromberger. Cambridge, MA: MIT Press.

Cowart, Wayne. 1997. Experimental syntax: applying objective methods to sentence judgments. Thousand Oaks, London, New Delhi: Sage Publications.

Devitt, Michael. 2006. Intuitions in linguistics. The British Journal for the Philosophy of Science 57: 481-513.

Duhem, Pierre. 1906/1954. The aim and structure of physical theory. Princeton: Princeton University Press. (The original publication in French in 1906, its original English translation in 1954, and its renewed edition in 1982. The page references are to the 1982 edition.)

Duhem, Pierre. 1996. Essays in the history and philosophy of science. Translated and edited, with introduction, by Roger Ariew and Peter Barker. Indiana: Hackett Publishing Company.

Feynman, Richard. 1965/1994. The character of physical law. New York: The Modern Library. (The page references are to the 1994 edition.)

Feynman, Richard. 1974. Cargo cult science. 1974 Caltech commencement address, reproduced in Feynman (1997, 338-346).

Feynman, Richard. 1997. Surely you're joking, Mr. Feynman! New York: Norton & Company.

Fitzgerald, Gareth. 2010. Linguistic intuitions. The British Journal for the Philosophy of Science 61: 123-160.

Fukui, Naoki. 1986. A theory of categories projection and its applications. Doctoral dissertation, MIT.

Hoji, Hajime. 1990. On the so-called overt pronouns in Japanese and Korean. In: Baek, Eung-In (ed.), Papers from the Seventh International Conference on Korean Linguistics. 61-78. Osaka: International Circle of Korean Linguistics & Osaka University of Economics and Law.

Hoji, Hajime. 1995. Demonstrative binding and Principle B. NELS 25: 255-271.

Hoji, Hajime. 2003. Falsifiability and repeatability in generative grammar: A case study of anaphora and scope dependency in Japanese. Lingua 113: 377-446.

Hoji, Hajime. 2006. *Otagai*. In: Ueyama, Ayumi (ed.), Theoretical and empirical studies of reference and anaphora — Toward the establishment of generative grammar as an empirical science. A report of the Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (B), Project No. 15320052, 126-138. Kyushu University (available at: http://www.gges.org/hoji/research/hp-papers.cgi).

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

Hoji, Hajime. 2010. Evaluating the lexical hypothesis about *otagai*. Linguistic Research 27.1: 65-119.

Kuroda, S.-Y. 1999. Bunpoo riron to tetsugaku teki sizen syugi [Grammatical theory and philosophical naturalism]. In: Chomsky, Noam and Kuroda, S.-Y. (eds.), Gengo to sikoo [Language and thought], 93-134 (translated by M. Ohishi). Tokyo: Syouhakusya.

Kuroda, S.-Y. 2008. Mathematics and generative grammar — Beyond explanatory adequacy and mathematical realism of language: A fable for Naoki Fukui. Sophia Linguistica 56: 1-36.

Lakatos, Imre. 1970. Falsification and methodology of scientific research programmes. In: Lakatos, Imre and Musgrave, Alan (eds.), Criticism and the growth of knowledge, 91-195. Cambridge: Cambridge University Press. Reprinted as Lakatos (1978: chapter 1).

Lakatos, Imre. 1973. Science and pseudoscience. Included in Lakatos (1978: 1-7) as "Introduction: Science and pseudoscience". (The page references are to Lakatos (1978). The transcript can be obtained at http://www.lse.ac.uk/collections/lakatos//Default.htm.)

Lakatos, Imre. 1978. The methodology of scientific research programmes. In: Worrall, John and Currie, Gregory (eds.), Philosophical papers volume 1. Cambridge: Cambridge University Press.
Lebeaux, David. 1983. A distributional difference between reciprocals and reflexives. Linguistic Inquiry 14: 723-30.
Meehl, E. Paul. 1967. Theory testing in psychology and physics: a methodological paradox. Philosophy of Science 34: 103-115. (Reproduced in Morrison and Henkel 1970/2007. The page reference is to Morrison and Henkel 1970/2007.)
Miyagawa, Shigeru and Arikawa, Koji. 2007. Locality in syntax and floating numeral quantifiers. Linguistic Inquiry 38: 645-670.
Morrison, E. Denton and Henkel, E. Ramon. (eds.) 1970/2007. The significance test controversy. New Brunswick, N.J.: Transaction Publishers.
Nakaya, Ukichiro. 1958. Kagaku no houhou [Method in science]. Tokyo: Iwanami.
Narita, Hiroki. 2010. The tension between explanatory and biological adequacy. A review of Naoki Fukui's (2006) Theoretical comparative syntax: Studies in macroparameters. Lingua 120: 1313-1323.
Newmeyer, J. Fredrick. 2008. A review of linguistic minimalism: origins, concepts, methods, and aims by Cedric Boeckx. Language 84: 387-395.
Nishigauchi, Taisuke. 1986. Quantification in syntax. Doctoral dissertation, University of Massachusetts.
Nishigauchi, Taisuke. 1992. Syntax of reciprocals in Japanese. Journal of East Asian Linguistics 1: 157-96.
Poincaré, Henri. 1952. Science and hypothesis. New York: Dover Publications. (The English translation of *La science et l'hypotheses* (1902).)
Popper, Karl. 1963. Science: Problems, aims, responsibilities. Federation Proceedings (Baltimore), Federations of American Societies of Experimental Biology 22.4: 961-972.
Reinhart, Tanya. 1983. Anaphora and semantic interpretation. Chicago: University of Chicago Press.
Saito, Mamoru. 1992. Long distance scrambling in Japanese. Journal of East Asian Linguistics 1: 69-118.
Saito, Mamoru. 2003. A derivational approach to the interpretation of scrambling chains. Lingua 113: 481-518.
Schütze, Carson. 1996. The empirical base of linguistics: Grammaticality judgments and linguistic methodology. Chicago: University of Chicago Press.
Sprouse, Jon. 2007. Continuous acceptability, categorical grammaticality, and experimental syntax, Biolinguistics 1: 123-134.
Townsend, J. David and Bever, G. Thomas. 2001. Sentence comprehension: The integration of habits and rules. Cambridge, MA: MIT Press.
Ueyama, Ayumi. 2010. Model of judgment making and hypotheses in generative grammar. In: Iwasaki, Shoichi; Hoji, Hajime; Clancy, Patricia; and Sohn, Sung-Ock (eds.), Japanese/Korean linguistics 17, 27-47. Stanford, CA: CSLI (available at: http://www.gges.org/hoji/research/hp-Ayumi.cgi).
Weinberg, Steven. 1992. Dreams of a final theory: The scientist's search for the ultimate laws of nature. New York: Pantheon Books.
Yoshimura, Noriko. 1992. Scrambling and anaphora in Japanese. Doctoral dissertation, University of Southern California.
Young, Charles. 2008. The great number crunch. Journal of Linguistics 44: 205–228.

Hajime Hoji
Department of Linguistics
University of Southern California
hoji@usc.edu