Title
Methodology in generative linguistics: scratching below the surface

Author
Johnson, Kent

Publication Date
2014-07-01

DOI
10.1016/j.langsci.2014.02.001

Language Sciences 44 (2014) 47–59

Contents lists available at ScienceDirect

Language Sciences

journal homepage: www.elsevier.com/locate/langsci

Methodology in generative linguistics: scratching below the surface

Kent Johnson *

Department of Logic and Philosophy of Science, University of California, Irvine, 3151 Social Science Plaza A, Irvine, CA 92697-5100, United States

A R T I C L E   I N F O

Article history:
Received 2 September 2013
Received in revised form 15 February 2014
Accepted 21 February 2014

Keywords:
Linguistics
Methodology
Data analysis
Theory confirmation

A B S T R A C T

This paper examines some methodological issues of contemporary linguistics from three general standpoints. A macro-level view sketches general, broad outlines of a technique. While this is the most common form of methodological discussion – by philosophers and linguists, including those of the "biolinguistic" persuasion – it is of limited use. In contrast, a micro-level perspective emphasizes those details that actually make a procedure do what it does. At this most important level of granularity, there has been very little development in generative linguistics. E.g., there are currently virtually no explicit means by which one might address a methodologically fundamental question about a data set such as "How much data is there?" A meso-level perspective allows for some aspects of a method to be compared with others. Inter alia, this perspective helps to identify narrower, more tractable methodological issues that currently need attention.

1. Introduction

> Psychology, beset with all of its conceptual difficulties, complexities, and logical challenges, has always harbored some of the most aggressive and ingenious adaptors of scientific developments from other fields.... Why this is so is patently obvious – the needs of the science are so great that the tendency to adopt anything that may help to design and execute experiments or that may provide a conceptual or metaphorical bridge to a satisfactory theory is overwhelming.
>
> (Uttal, 2003, 45)

Since its inception over a half-century ago, generative linguistics has aimed to characterize the distinctively linguistic components of the human mind. During much of this period, linguists have developed increasingly sophisticated theories with two main goals in mind: to explain both the remarkable variety of linguistic phenomena, and also how such phenomena can be appropriately realized by normal humans in normal conditions. Much of this theorizing has coalesced around theories of the *faculty of human language* (FHL), a hypothesized cognitive structure that links these two goals. Over the years, theories of FHL have become increasingly sophisticated. The relatively recent rise of the biolinguistic program has continued this tradition by suggesting additional constraints on theories of FHL that aid its biological plausibility.

In what follows, I consider some key aspects of the *methods* by which such theories are derived, confirmed, criticized, etc. In many respects, nothing could be more ordinary. Nearly every other area of empirical research has many journals, books, societies, faculty lines, etc. devoted to such matters. E.g., consider such journals as *Biometrics, Psychometrika, The Journal of*

* Corresponding author. Tel.: +1 (949) 824 8968, Office: SST 755.
E-mail address: johnsonk@uci.edu
URL: http://www.lps.uci.edu/~johnsonk

http://dx.doi.org/10.1016/j.langsci.2014.02.001
0388-0001/

Mathematical Psychology, Psychological Methods, Econometrica, Journal of Chemometrics, Journal of Chemical Information and Modeling, Chemometrics and Intelligent Laboratory Systems. Cf. Section 6 for discussion and details. However, while research into the structure of FHL has flourished, the manner by which such research is conducted has received virtually no attention, at least in the central component of linguistics this paper addresses; cf. (1). That is, while the list of journals above could be extended, it still wouldn't contain a "*Linguometrica*". But the methodological needs of linguistics are at least as great as those of the other fields of empirical research.

This paper has two main theses. First, I argue that generative linguistics should follow the age-old scientific trend of vigorously analyzing and improving explicit methods appropriate to the discipline. Such a project can be entirely complementary with the primary study of FHL, and the two (to the extent they differ) would naturally be pursued simultaneously and symbiotically. Second, I also maintain that one of the reasons this hasn't happened is that there has already been a long tradition of methodological discussion in linguistics. However, this discussion has taken place at a relatively superficial level. Such discussions, I argue, can serve some important purposes, but they are no replacement for the level of detail required for the needed methodological advancements.

A particular method or procedure can often be viewed in many ways; this paper considers three such rough-and-ready categories: a very general "macro-level" perspective; a quite detailed "micro-level" one; and a middling "meso-level" view. These categories are purely heuristic, and do not indicate any sharp boundaries; my use of *methodology* and like terms is similarly loose and general.[1] Despite their informality, these notions still point to some important distinctions. The most basic message about these views is that they are not on an equal par. Macro- and meso-level views of methods have historically received much more attention in the relevant literature, partly because they supply more readily accessible presentations. But the missing details are crucial. Indeed, they determine whether, why, and in what respects the macro- and meso-level claims are true.

Across the sciences, the emphasis on quantitative micro-level details is absolutely standard. Indeed, the field-specific study of a field's methods are often the topic of whole subfields (cf. below; Uttal, 2003; Danks and Eberhardt, 2009 for some discussion). In like fashion, I think that the need for attention to micro-level methodological details in linguistics is particularly important. Below I discuss the need for three kinds of micro-level developments. (i) Most generally and importantly, linguistics is in desperate need of an explicit quantitative theory of the uncertainties inherent in theorizing about FHL. Such a theory itself should have various features, such as (ii) the ability to aid in the assessment of a data sets' evidential bearing on a given theory. In particular, it should help to replace the false but familiar view that the "best" available theory should be "tentatively accepted as true" with a much more fine-grained analysis. But these bigger methodological goals can themselves be met only by attention to myriad more humble matters. For example, (iii) no serious advancement in linguistic methods can be made without a better sense of such basic characteristics of a body of data as its size. How much data is there in a given study – a little? a lot? More than 15 observations? As fundamental as this issue is, how one might even *begin* to go about addressing it is itself virtually never addressed in the core methodological approaches to generative grammar. It is nowhere addressed in, say, (1):

(1) (Baker, 1988; Boeckx, 2006; Chomsky, 1981, 1995; Grimshaw, 1990; Hale and Keyser, 1986, 1987; Hornstein, 2009; Kayne, 1994; Williams, 1980, 1983; Baker, 2001; Boeckx and Piattelli-Palmarini, 2005; Chomsky, 2000, 2002; Walenski and Ullman, 2005; Chomsky, 2005, 2007; DiSciullo and Boeckx, 2011; Epstein, 2007; Jackendoff, 2002, 2011)

(1) is just a sample; it could be increased by orders of magnitude by merely examining a few issues of, say, *Linguistic Inquiry, Natural Language and Linguistic Theory, Biolinguistics*, etc. Moreover, this fundamental matter is more challenging in linguistics than simply counting up participants in a study; rather it appears similar to some of the issues of data reduction in multivariate data analysis. But this additional complexity only strengthens the need for explicit methods, to say nothing of the great many more (and more difficult) matters lying at the core of empirical research, linguistic and otherwise.

Before beginning, a few caveats are in order. This paper focuses on those methods most commonly employed in Chomskyan generative grammar, and the term *linguistics* here covers only this core component of the discipline. E.g., I have in mind general technical works, theoretical discussions, many biolinguistic discussions, etc. such as those in (1). However, I do not directly address such areas as corpus linguistics, computational linguistics, neurolinguistics, psycholinguistics, and many other types of research [Schütze (1996) supplies a much broader overview of various linguistic methods].

Additionally, this paper concerns how given evidence is taken to bear on the construction and confirmation of linguistic theories, especially components and computational processes of FHL. In particular, judgments of grammaticality (inter alia) are assumed to be correct, and no questions of their reliability, validity, etc. arise here. Thus, this paper contrasts with much current methodological research concerning how linguistic data *should* be generated. E.g., some authors defend the standard methods of evidence collection, where a single person, often the linguist herself, confronts one or two instances of a construction, and makes a judgment about the status of the construction, e.g., (Gross and Culbertson, 2011; Phillips, 2009; Culicover and Jackendoff, 2010); also (Sprouse and Almeida, 2010, 2012) use statistical methods to validate these

[1] Terms like *method* etc., only gesture at how a typical practitioner goes about doing what she does, perhaps with reference to her various goals and the materials available to her. Similarly, I will often write as if a research area had a single, consistent, stable, universally applicable methodology. This is importantly false, as discussed in detail below. But since the task at hand turns on the details of concrete methodological examples – as opposed to the entire methodological toolkit that various researchers in a given field (however this latter notion is spelled out) employ – this literary convenience will do no harm.

non-statistical methods. Others, e.g. (Gibson and Fedorenko, 2010, 2013) argue that linguistic evidence should be produced by methods more familiar in the cognitive sciences, where multiple naive subjects deliver reports about multiple instances of a construction, and the resulting body of data is statistically summarized and analyzed. These issues, while interesting and important, are orthogonal to the present paper. Intuitively, we can think of these latter issues as addressing the question "How do we collect good evidence?"; this paper asks "How do we understand the relationship between the available evidence at the theories it is intended to bear on?"[2]

This paper is organized as follows. Section 2 sets out some background issues, especially regarding the fundamental object of study, FHL. Section 3 illustrates macro- meso- and micro-level views of a well-understood and fully explicit technique of data analysis, principal component analysis. The next three sections discuss linguistic methods from each of these perspectives. Section 4 considers some general methodological contours of linguistic theorizing. While such broad perspectives have their place, here they are brought out primarily to avoid confusing them with the methodologically more crucial details. Section 5 uses a couple examples to begin to chip away at the vast gulf between a data set and its ultimate application to a theory about some bit of FHL. Section 6 uses some insights from other areas to reinforce the ideas of Section 5. It also examines how some general techniques of latent variable analysis might suggest some means by which further methodological advancements in linguistics might be made. Section 7 concludes the paper.

2. The background: hypotheses and assumptions

This section characterizes some relevant background hypotheses, assumptions, issues, etc. that help determine the outlines of linguistic theories. Ultimately, I am concerned not so much with these theories themselves, as with the methods for assessing them. Of course, the nature of the particular hypotheses (and, typically, its idealized representation) play a key role in determining the appropriate methods of assessment. Nevertheless, a reasonably sharp distinction can still be maintained between, on the one hand, the theoretical claims of interest, and on the other hand, the methods used to (dis-)confirm them. In scientific terms, this distinction is commonly made as that between a hypothesis (e.g., a null hypothesis of no difference in the means of the treatment and control population) and the procedures by which it is assessed (e.g., by a two-sided *t*-test).

Broadly speaking, the main underlying assumption of generative linguistics is that at some scientifically interesting level(s) of (psychological, neurological, biological) characterization, there exists a "mental organ", FHL, which is responsible for normal humans' (in normal environments) abilities to acquire and speak a language (where this latter term is used somewhat loosely here; cf. Chomsky, 1995, 1). Normally, the hypothesis goes, a child's FHL attains some particular state, an *I-language*. This I-language algorithmically produces certain kinds of cognitive data structures, *I-expressions*, which do not necessarily correspond to any pre-theoretic notion of linguistic expressions. Rather, I-expressions are assumed to interact with various other components of a person's psychological (and neurological, biological, etc.) makeup so as to ultimately result in key aspects of the comprehension and production of natural language. Much more than just FHL is involved in a person's ability to use her language; however, the guiding hypothesis is that there is scientific merit in isolating something like FHL that plays something like the role just described. Understanding FHL and its roles is the central goal of linguistics.

The basic ideas just sketched will be revisited later in Section 6. For now, it suffices to observes that it fits Fitch's comment that "[t]he basic hope of explaining complex surface phenomena by virtue of simpler but more abstract principles underlies all science" (Fitch, 2010, 383). Similarly, it fits with a perennial view of psychology: "For Wundt, James and others psychology was the science of mental life, or of conscious mental states. Behavior might provide a form of evidence, but there was no thought psychology was aimed at predicting behavior" (Hatfield, 2002, 213–214). Moreover, this view also supports a common sentiment that linguistics is not simply a "science of patterns" of grammaticality and ungrammaticality. Although some of the *evidence* might occur as grammaticality judgments of various sorts, such data serves to uncover latent cognitive structures.

This story also presents us with the outlines of a theory, one that can be filled out in various of its details and then tested, (dis)confirmed, improved, integrated with other "neighboring" theories and so on. Notice also that it takes us from our primary interest in the nature of FHL to particular I-languages, and then to particular I-expressions, and ultimately to particular bits of overt linguistic behavior, broadly construed. One important, and very common, strategy for investigation follows this sequence in reverse order. We have access to speakers' linguistic behavior, which may be in the form of a judgment that, say, it is acceptable to use (2a) but not (2b) to inquire into the identity of the person who, together with Sue, carried the piano upstairs:

(2) a. Who did Sue carry the piano upstairs with?
b. * Who did and Sue carry the piano upstairs together?

Such facts can be used as evidence about the nature and role of FHL. Of course, evidence regarding FHL may, in principle, come from anywhere; e.g., other areas of linguistics, psychology, biology, anthropology, mathematics, etc. (e.g., Fitch,

[2] Of course, this characterization assumes a certain perspective about what counts as "evidence". For present purposes, we can regard it as the final assessments about grammaticality of various expressions, whether they were produced from a single expert judgment, or as the amalgamation of many judgments, perhaps of varying constructions.

2010; Chomsky, 2005; Hauser et al., 2002). Similarly, confirmation needn't occur only in the straightforward path just described.[3] The present focus, however, is only on the type of evidence just exemplified, which, as it happens, is currently by far the most widely used type.

The issues considered below should be carefully distinguished from two other matters. The first concerns the formal structures involved in linguistic theorizing. Consider e.g., the familiar labeled tree diagrams (and their linear counterparts) that characterize the relevant structure of candidate I-expressions. Such diagrams often make fully transparent how a given I-language might algorithmically generate certain structures, while failing to generate others. The precision and clarity of this aspect of linguistic theorizing is nearly perfect, and in that sense are beyond reproach. However, these are features of the hypotheses themselves. It is another matter entirely how such a theory, or collections of rival theories, are justified, (dis)confirmed, etc. in the light of evidence and existing theory.

The second issue concerns the most widely discussed philosophical issue about linguistic theories; viz., our knowledge of language. Many philosophers and linguists have weighed in on the question "in what our knowledge of language consists" (some well-known positions include Chomsky, 1986, George, 1989a, Dummett, 1993, Higginbotham, 1989). Much of this discussion traces back to Quine's earlier charge that the most that can be said of any given grammar is that it "fits" the speaker, in the sense that the speaker "conforms" to it (v. O. Quine, 1972, 442; cf. Johnson, forthcoming). Quine's thought here is that there will always be infinitely many other "extensionally equivalent" grammars – i.e., grammars that produce the very same set of expressions as the original one. Thus, he argued, if there is no way to tease them apart empirically, none of them can be distinguished from the others, and regarded as truly characterizing FHL and its workings. That is, no grammar can be held as "guiding" a speaker's behavior, in Chomsky's sense (Chomsky, 1986) of accurately characterizing the psychological (etc.) components that constitute the speaker's linguistic competence.[4] Quine saw this as a permanent and principled problem. He held that the available data don't merely *underdetermine* which particular grammar should be accepted, but that they could never do this. Rather, there are always multiple grammars that accommodate the data, so it is *indeterminate* which one is "correct".

Knowledge of language, while important, should not be confused with the present issues. The former concerns features of the objects being researched: human beings *qua* language possessors, and how they are related to their grammars. As such, this issue concerns the roles of a linguistic theory and its bearing on various philosophical issues. But it does not bear directly on the practices by which linguistic theories are constructed and assessed.

3. Macro- meso- and micro-level views: illustrations

In this section, I sketch the rough-and-ready categories, mentioned in Section 1, of perspectives on a given method that one might adopt. In order to later expose some of the more relevant details of linguistic theorizing, I will here use an example that has been thoroughly analyzed, principal components analysis.

To get things started, consider a study of the optic system of goldfish involving 10 different receptor cone cells, with their absorption rates recorded for 7 different light wavelengths, given in (4), truncated from Weller and Romney (1990, 9).

(4)

Violet	*Indigo*	*Blue*	*Green*	*Yellow*	*Orange*	*Red*
147	153	89	4	0	0	0
153	154	110	24	23	17	0
145	152	125	0	0	0	0
99	101	122	153	93	44	0
46	85	103	148	116	75	26
73	78	85	154	109	57	0
14	2	46	106	137	92	45
44	65	77	102	151	154	120
87	59	58	79	139	163	146
60	27	23	72	136	144	111

[3] The development and confirmation of a theory is always a holistic affair, often proceeding in a "bootstrap" fashion (Glymour, 1980). Some well-confirmed aspect(s) of FHL or some particular I-language(s), say, may impact how we understand some more "lower-level" elements, such as the structure of some I-expression(s). Likewise, the linguistic data presented in journal articles, say, is itself partly the product of existing theory. E.g., most linguists treat the data in (3) as straightforwardly generated by a typical English speaker's I-language, even though such speakers may treat them as robustly unacceptable

(3) a. The horse raced past the barn fell.
b. The gnat the cat the dog the man owned chased ate died.

The strong linguistic and psycholinguistic background theory suggests the unacceptability is largely due to details of parsing. Such theoretical "preprocessing" of data is not uncommon; frequently an analysis begins by examining the data for errata, outliers, or other inappropriate features of the data that should be corrected before the analysis proceeds.

[4] For Quine, a rule *guides* a speaker only if the speaker "knows the rule and can state it"; i.e., she "*observes* the rule" (v. O. Quine, 1972, 442). Chomsky's discussion uses a notion of guidance more like the one characterized above. In particular, the grammar that guides the speaker, Chomsky argues, can be held to be causally efficacious in ways that extensionally equivalent ones cannot. [George's valuable distinctions between various uses of the term *grammar* are of much use here (George, 1989a).] This discrepancy is purely terminological, though; Chomsky's argument is designed to directly address Quine's concerns.

Many of these cones' absorption profiles are fairly similar. We might wish to use the data in (4) to drive an estimate of how many different types of cones there really are. One of the most well-known methods for exploring such matters is *principal components analysis* (PCA) (e.g., Jolliffe, 2010). There are many different ways to describe what PCA does; here are three.

(I) Broadly speaking, PCA can help uncover the underlying latent structure of a set of correlated measurements. Using (4), PCA reveals that 98.1% of the information in it can be captured with just three distinct types of underlying structures. This suggests that the goldfish eye may contain just three types of cone (like the human eye). In short, a PCA may help identify a few latent constructs that are primarily responsible for a much larger body of data. Further research might additionally help supply an empirical interpretation of these features of the data that PCA reveals.[5] Looking for structure *within* a set of variables can then be seen to contrast with, say, regression analysis, whereby practitioners often try to obtain a "best" *prediction* of the value of one variable, given the values of a set of other variables.

(II) PCA can also be described as treating each of the 10 cones as a distinct geometric axis, so that each of the seven wavelengths constitute a point in a 10-dimensional space. The PCA itself is an orthogonal rotation of these 10 axes, so that the kth "new" axis captures as much variation in the data (i.e., the wavelength absorptions) that has not been captured by the previous $k-1$ "new" axes (e.g., Dunteman, 1989; Wickens, 1995). To visualize PCA, imagine a situation where the absorptions are measured for only three cones, and the data is plotted in a three dimensional "cube", where the x, y, and z axes correspond to the three vectors of measurements. Suppose also that the data form a cloud, centered at the origin of x, y, and z, in roughly the shape of a flattened football floating at some angle. The first new axis will pass through the long main part of the football, the second through the wide, flattened part, and the third through the thin remaining part. Representing the data using only the first two new axes – or any two axes that span the same subspace – you would capture most of the original information, since there is little remaining variation for the third axis to represent.

(III) Finally, PCA can be mathematically exhaustively characterized. This involves e.g., identifying the kth PC and its properties (which were formerly given with looser terms like "representing", "items responsible for", "capturing the most remaining information", etc.) with mathematically precise structures and properties. In particular: the kth PC is a linear combination of the 10 original variables whose coefficients are the scaled elements of a unit-length eigenvector with the kth largest eigenvalue of the 10×10 covariance matrix of the original measurements. It can be shown that this PC (i) is orthogonal, according to the euclidean inner product, to the $k-1$ dimensional subspace determined by the previous k PCs; and (ii) determines a 1-dimensional subspace such that the sum of the euclidean orthogonal projections of the 7 original measurements onto it is no smaller than any other 1-dimensional subspace satisfying (i). (Obviously, since only 7 wavelengths are measured, only the first six PCs will identify proper subspaces.) [Cf. e.g. (Schott, 2005, 104–110), (Magnus and Neudecker, 2007; Jolliffe, 2010, chap. 17) for details.]

These three descriptions illustrate macro-, meso-, and micro-level perspectives on PCA. A brief bit about each perspective follows.

The *macro-level* view in (I) emphasizes the "big picture" themes, such as the broad patterns by which PCA operates, the typical goals of using it, etc. By abstracting away from many of the particular details of the technique, we can get a better sense of the technique's intended purpose, and how it is supposed to fit into the larger enterprise at hand. Here it is natural to set aside many real-life complications, and e.g. simply assume that the method is being used correctly and successfully, the inputs to the technique were appropriate, etc. This generic purview of a given method is often helpful for considering general topics in the philosophy of science, such as underdetermination and indeterminacy, along with scientific explanation, scientific realism, theory confirmation, reductionism, eliminativism, etc. Importantly, one can discuss the method of PCA at this broad level without understanding its mathematical structure or how it is actually implemented.

The *meso-level* view in (II) adds some more detail to (I), so as to display its relationships to other areas. E.g., (II) characterizes PCA as capturing information in a geometric fashion. As such, (II) suggests that PCA might be better understood/improved/modified, etc. with geometric methods.[6] This middling perspective involves enough information about the details of how the method works to make connections with those of other techniques that might superficially seem rather different.

The *micro-level* view in (III) presents the real nuts and bolts of PCA While (I) and (II) can serve valuable purposes, their justification depends entirely on the details articulated in (III). After all, it's one thing to describe a technique as capturing as much remaining information as possible on the kth new axis, and it's quite another for it to actually do so. That is, there is an obvious sense in which the micro-level details are most important.

Importantly, micro-level subtasks like determining the coordinates of the first k PCs, along with establishing their relevant optimizing properties, are not simple. That is, it takes explicit calculations and background mathematical theory to determine that 98.1% of the total (row) variance is projectable onto a three-dimensional subspace of the seven-dimensional space spanned by (4)'s row vectors. This is not revealed merely by visual inspection of (4).

As generative linguistics is currently practiced, essentially no attention has been paid to the development and inner workings of the methods by which such theorizing occurs. Instead, some loose macro-level discussion of linguistic methodology has obscured the fact that there has been no careful micro-level study or development of the methods by which linguistic

[5] As always, there are no panaceas here (e.g., Fabrigar et al., 1999; Gould, 1996).

[6] As a matter of fact, the intellectual origins of this technique of data analysis were geometric (Pearson, 1901). Moreover, it eventually came to be seen as subsumed under another, more general technique, the singular value decomposition, in part via the latter's own geometric interpretation (Eckart and Young, 1939; Weller and Romney, 1990; Horn and Johnson, 1985, 1990).

activity occurs. That, in part, is why linguistic methods, in contrast to an application of PCA e.g., are never explicitly articulated. Consequently, the appropriateness of these methods, along with their inferential characteristics, cannot be explicitly studied. (As discussed above and below, this is extremely unusual for an empirical discipline.)

4. A macro-level of analysis

In the next three sections (Sections 4–6), I discuss some roles that macro-, micro-, and meso-levels views of linguistic theorizing can play. Although they are each potentially useful, it is nevertheless crucial not to conflate the different sorts of goals involved. Most importantly, informal general macro-level discussions of theoretical "models" and "tests" of hypotheses are no substitute for precise micro-level treatments of these notions, including an explicit theory of uncertainty.

As we've seen, a macro-level description of some technique brings out its general structure and workings, its users' typical goals, etc. Such characterizations tend to be widely accessible, and so are quite common. Here is a simple, albeit extreme, example from linguistics and philosophy.

Above in Section 2, I briefly mentioned Quine's argument that a grammar (i.e., a formal theory of an I-language, or of FHL) could at best "fit" the empirical facts, but couldn't "guide" a speaker. Chomsky's reply to Quine is subtle, but for present purposes, the following sketch will do. Buttressed by numerous examples from the history of science, Chomsky holds that it is a normal part of everyday scientific inquiry to regard the (possibly unobserved) structure posited by the best available theory of some phenomena as the true causal and explanatory structure that accounts for the evidence under study (e.g., Chomsky, 1986, 1995). Similarly for linguistics, if we are (tentatively) entitled to regard a theory of the human faculty of language as true, then we are (tentatively) entitled to regard the structure it posits as characterizing how a speaker's linguistic behavior is *guided* by that faculty; cf. Section 6 for more subtlety on this. Over the decades, Chomsky has reiterated this sentiment in many different places (e.g., the papers in Chomsky, 2000; cf. also Chomsky, 2005, Chomsky, 2007). E.g.:

> [W]e propose certain principles, parameters, representation, modes of computation, and so forth, and we seek to explain the facts in these terms, taking this account tentatively to express the truth about the language faculty. Although there are distinctions...we do the same when we attribute a certain structure to a molecule of water or benzene, to a cell of the body, or to the sun to account for light emissions, or to a machine of some sort. If our best theory accounts for Jones's behavior by invoking these rules and other elements, we conclude that they enter into Jones's behavior and guide it, that they play a "causal role" in the sense of this discussion.
>
> (Chomsky, 1986, 249)

(Talk of causal roles must be taken with some delicacy here; Chomsky has in mind the kind of role appropriate to a computational level of description of FHL in something like Marr's sense; Marr, 1982, 22–29; cf. Chomsky, 2000, 23, 158–159; Chomsky, 2002, 74; Chomsky and McGilvray, 2012, 213.)

From the present perspective, then, Chomsky's view can be summed up as follows: *when used successfully*, the methods of linguistics supply reason for accepting the cognitive structures it posits to explain the available linguistic evidence. This claim is justified, so the view goes, because, at the present level of magnification, the methods of linguistics are identical (or very nearly so) to the methods of the other sciences, which similarly provide justification for accepting their posits. That is, Quine's in-principle problem of indeterminacy is addressed by abstracting away from many details and assuming that the linguistic theory has been appropriately successful according to the ordinary standards of other areas of empirical research. In such a case,

> We proceed in practice by taking a realist stance toward theoretical discourse. Although some may feel that this stance is somehow illegitimate, that it goes beyond the evidence, the issues are not crucially different from those that arise in any intellectual work of significance. The significant questions have to do with the persuasiveness and explanatory power of the theories and *the quality and scale of the evidence bearing on them*.
>
> (Chomsky, 1986, 252, emphasis added)

To make this comparison between linguistics and "any intellectual work of significance", one must necessarily abstract away from the particular methods used by linguistics and these other works of significance, and attend only to the general logical relationships between evidence and theory. This is a natural sort of in-principle reply to an in-principle problem.

In some sense, the reply to Quine sketched above is especially appropriate, since Quine doesn't – and shouldn't, given his aims – ever suggest that a proposed explanation of, say, the phenomena illustrated by (2) suffers because the evidence is merely not strong enough. Quine's position, after all, was much stronger than this.[7]

Leaving Quine's philosophical challenges aside, the remaining matter of the precise nature of the evidence and its relation to some given theory is still crucial. Indeed, this general issue is of absolutely central importance. Chomsky recognizes as much in his adversion to the "quality and scale of the evidence".[8] I take it that this refers to the strength of evidence, and the amount of it. In your standard experimental design, the underlying model allows for these two features to be mathematically represented and combined into a single index, such as a p value, which is often interpreted as the strength

[7] Quine suffered from what Wilson dubs *theory-T syndrome*; the latter occurs when one abstracts away from so many details that the resulting schema of a theory is too devoid of content to be of much service to science or philosophy (Wilson, 2006, 127).

[8] Cf. also the comment that "it is not in doubt that evidence has been provided, *however one may assess its strength* (Chomsky, 1986, 248, emphasis added).

of evidence. Even in the simplest case, such assessments cannot – and *should not* – be done "on the fly", or in one's head; rather, they require explicit calculations. (Indeed, it is to the credit of such methods that they are so explicit that they can be trenchantly critiqued, improved, formally compared with alternatives, etc. [e.g., Cox, 1958; Cohen, 1994; Royall, 1997; Howson and Urbach, 2006].)

Issues regarding assessing evidence become central when we step back from the rarified question whether some particular theory of FHL is *justifiable*, even in principle, and return to the more ordinary and pressing question whether a particular bit of it is in fact *justified*, and if so, exactly how it is so. The latter questions are ordinarily decomposed into myriad smaller matters – as we'll see, "quality" and "scale" are themselves placeholders for narrower, tougher, and usually more explicitly tractable issues, not ones to be assessed simply by thoughtful reflection. This is an area where linguistics has not joined the other empirical fields in expending more energy developing more, and more explicit, methods for understanding the relevant data and its collective bearings on various theories.

In sum, macro-level views of some bit of methodology can be useful for big-picture presentations that are accessible to non-specialists (cf. Carey, 2009, 43–44 for a particularly clear example of this). They may also at times serve some purpose in philosophical discussions. However, such descriptions themselves are of limited justificatory import. Instead, they are merely convenient placeholders for the much more complex, and virtually always mathematically involved, processes that they stand in for. Thus, from the present perspective, the issues scouted here can be joined with those background issues from Section 2 as matters that should not distract us from the ones that do the real epistemic work.

5. A micro-level of analysis

Above in Section 3, we saw that one can describe a procedure like PCA in various ways. PCA is a well-understood and fully explicit technique. I.e., the micro-level details that support the breezier descriptions are well-understood. Thus, the reason why a given PC captures as much of the remaining variance as possible, can be captured in a mathematical proof. In this sense, it illustrates a goal or standard to which linguistic theorizing should strive to attain.

In this section, I consider a concrete case of how one might get started trying to realize such a goal of methodological explicitness within linguistics. In particular, I consider the simple but vital micro-level question: *how much data was used in a given linguistic project?* (A "project" can be individuated in any reasonable way, such as a paper or collection of papers that you regard as unified around some theme.)

In general, the relative size of a body of data under consideration is crucial to its analysis, especially in those areas (of which linguistics is clearly one) where there is much uncertainty and/or complexity in relating data to theory. Indeed, it is hard to think of a more fundamental characteristic of a body of data than its size. This is readily apparent even in the simple case of flipping a coin to determine if it is biased. The value of knowing that the coin came up heads 60% of the time depends heavily on the total number n of flips. If $n = 10$, the evidence is weak ($p = .53$); if $n = 100$, it is stronger ($p = .05$). Similarly, all of the familiar tools of data analysis, modeling, and hypothesis formation and testing depend on the size a data set. These tools extend from basic methods like point estimation and confidence intervals, all the way to identification and optimization of latent structure, testing whether that latter fits the data well, the amount of work individual bits of it do, etc.

As even the coin-flipping case shows, understanding such basic characteristics of a data set as its size is crucial to any assessment of how the data bear on a given hypothesis. The very simplest model of coin flipping involves the binomial distribution; without access to the size of the data set being used, the key formula would reduce to such nonsense as

(5) $\binom{Many}{Some}.5^{Some}(1-.5)^{Many-Some}.$

Most readers will find (5) jarring, which testifies to just how little can be done without more explicit information about the collective body of data. The case of linguistics is much more complex, but this only increases the need for greater clarity about such basic matters. Remarkably, though, such issues have gone basically ignored in linguistics. Nothing of the sort is even mentioned in the works in (1). (The topic of whether a child learning language might (unconsciously, of course) exploit statistical regularities and other principles of data analysis is sometimes acknowledged (e.g., Chomsky, 2005, 6). However, this is beside the present issue concerning the explicit analysis of a data set's characteristics by the *linguist*.) Worse yet, the case of linguistic data sets is considerably more difficult than simply counting coin flips.

Notice that issues of size already play a role in assessments of various sorts in linguistics. They appear whenever one judges that a journal article, research project, presentation, etc. uses "a lot" of data, "not enough" data, a "wide range" of "diverse" data, "more" data than a rival hypothesis is founded on, etc. Such comments are clearly meant to summarize certain aspects of the evidence, and to be one component of an overall normative judgment regarding the current evidential support for the theory.

Issues such as the overall magnitude of a body of linguistic data are only small pieces of the global assessment of the data's bearing on a given hypothesis about FHL. However, such issues also hold out correspondingly more hope of being tractable, and issues of tractability are highly relevant here. After all, the complexity of the general task is enormous. E.g., in linguistics the basic empirical evidence about FHL is often some unknown combination of outputs from a speaker's I-language combined with additional components of "noise". (Indeed, it is almost definitional that the more abstract and latent the underlying phenomena of interest is posited to be, the more irrelevant noise the actual observable data contain.) Moreover, additional uncertainty is brought in from the fact that actual linguistic theories try to make sense out of only tiny bits of an enormously complex interrelated whole (this point is familiar in the philosophy of science (e.g., Suppes, 1962; Trout, 1998; Mayo, 1996). In short, multiple sources of uncertainty conspire to make it notoriously difficult to confidently and precisely adjudicate between the many competing theories that are found in the literature. (I have in mind the kinds of debates one finds in the pages of, say, *Linguistic Inquiry, NLLT, Syntax*, etc., such as whether control should be eliminated in favor of movement, whether the current minimalist notion of a phase should be adjusted, what thematic roles have to do with the middle construction, etc.) In general, the more uncertainty there is, the more that large-scale assessments need to be analyzed into smaller, hopefully more tractable, bits.

Ultimately, vague talk of "a lot" of data must be replaced with specific talk of particular quantities. To see what I have in mind, take a few pages from *The Minimalist Program*, say, and try to answer the question of how much data was used. More than 15 data points? More than the next five pages of the chapter? If so, how much more?

These questions are not simple. The problem is that linguistic data can easily be wholly or partially redundant. E.g., if you use the data in (2) in a study of raising phenomena, (6) "adds" little relevant further data:

(6) a_1. Who did Pete carry the piano upstairs with?
a_2. Who did Sue carry the sofa upstairs with?
a_3. Who did Pete carry the big red box upstairs with?
⋮
b_1. * Who did and Pete carry the piano upstairs together?
b_2. * Who did and Sue carry the sofa upstairs together?
b_3. * Who did and Pete carry the big red box upstairs together?
⋮

Although (6) could be made countably infinite in size, it is sufficiently redundant with (2) in the relevant respects that it tells us little more about (the relevant aspects of) the empirical world than we already knew from (2).

Elsewhere (Johnson, 2011), I examined this kind of redundancy problem for estimating the size of a linguistic data set. In attempting to develop an explicit method for accounting for redundancy, I ended up proving a theorem that nothing like it would work in general. In fact, the impossibility result proved to be quite robust in a wide variety of ways. This demonstrates that dealing with redundancy is a highly nontrivial matter. For example, a researcher may wish to assess the evidence for a given theory in part by adjusting for the relevant correlations between the various linguistic constructions in the data at hand. However, (i) there is little reason to think that this can be done accurately at an implicit, informal level, as part of the researcher's overall "expert judgment"; cf. (Johnson, 2009) for extended discussion. In the present case, this should be clear on the face of things[9]; however, there is also a vast literature on the unreliability of expert judgments in tasks that are typically much simpler and clearer than the present type. Indeed, such informal judgments famously function more poorly than their users believe (e.g., Dawes, 1979; Arkes, 2003; Koehler and Harvey, 2007; Bishop and Trout, 2002, 2005; Johnson, 2007; Meehl, 1954, 1986; Grove and Meehl, 1996). (ii) No other field of empirical research is so utterly dependent on such implicit practices; cf. below Section 6. Moreover, as Johnson (2009) discusses, there is currently no known way to assess how accurate, reliable, etc. such implicit judgements are linguistics. A fortiori, the accuracy of such practices have in fact never been assessed. Finally, (iii) the result in Johnson (2011) demonstrates that some goals which might be thought to be attained by implicit judgement are in fact mathematically impossible.

In contrast, many other areas – from psychology and the social sciences to biology, chemistry, computer science, and physics – have dealt explicitly with issues of redundancy in data. As a result, a wide variety of formal tools have been developed and refined (e.g., Jolliffe, 2010; Basilevsky, 1994; Weller and Romney, 1990; Greenacre, 2007; Lazarsfeld and Henry, 1968; Hagenaars and McCutcheon, 2002). The fact that so many other areas of research have dealt with somewhat similar issues suggests that redundancy counts as a methodological *problem*, in Chomsky's sense, and not a mystery that we have no idea how to make progress on Chomsky (1976, 281). This is not to say that the problem is easily solved. (E.g., the theorem mentioned rules out most of the most common techniques, such as PCA, for the appropriate sort of data reduction.)

Unlike the current implicit global judgment, the micro-level details of explicit methods should allow for precise and principled answers to such questions as:

(7) In a given linguistic project, to the nearest 5 (0, 5, 10, 15, 20, ...), how much data (e.g., how many non-redundant expressions) were used? How was this assessed? What means were used to correct for redundancy? What ensures that these methods actually work as claimed, such as a theorem to this effect?

These questions are about as basic as they come, and are standard fare across a great many disciplines. Thus, they are good ones to start trying to answer – and good initial benchmarks of where work needs to be done, to the extent their answers are unknown. It's plausible that a heightened understanding of the multivariate features of linguistic data would eventually

[9] E.g., in the negative result in Johnson (2011), the problem fundamentally involves the existence of negative eigenvalues of a matrix that is derived from the original collection of linguistic data. To put it mildly, the existence of such eigenvalues is not something that one can see by simply examining the data itself, and checking whether one has an intuitive feel for how it all works out.

replace (7) with more sophisticated ones. But such a suspicion only further motivates the need for increased attention to the development and refinement of explicit methods for dealing with linguistic data.

Obviously, the challenges at issue don't end with estimating a data set's size. Rather, there are a great many characteristics of a data set and the inferential tasks set to them that should be made as explicit as possible.[10] Moreover, various macro-level desiderata such as the (tentative) acceptance of some bit of theory of FHL should be built up as much as possible out of the much smaller micro-level details we've been discussing.

To see this last point, consider a familiar attitude about the special role accorded to the overall "best theory" of some phenomenon. The idea is that such a theory has earned the right to have its hypothesized component processes and structures taken seriously and investigated as such. E.g., regarding linguistics, Chomsky writes

> I cannot see that anything is involved in attributing causal efficacy to rules beyond the claim that these rules are constituent elements of the states postulated in an explanatory theory of behavior and enter into our best account of this behavior.
>
> (Chomsky, 1986, 253)

Though familiar, this idea is misleading. Theory acceptance is more complicated; e.g., sometimes our best theories simply aren't good enough. In such cases, rather than tentatively regarding our best theory as true, we instead draw weaker conclusions, or perhaps no conclusions at all. A textbook case occurs when using a random sample to estimate the mean μ of a normally distributed population. Our best theory of the value of the unknown constant is given by the maximum likelihood estimate of it. But this best theory must also be checked against, say, a confidence interval in order to get some grasp of the uncertainty of the estimate. E.g., a point estimate of 550 nm for an energy wavelength that has a (95%) margin of error of ± 2 nm is much more informative than the same point estimate with a (95%) margin of error of ± 300 nm. Indeed, depending on the level of precision required, in the latter case, it might be natural not to accept any theory, best or otherwise.[11]

The situation for linguistics is similar. One might identify a unique best theory (of control, anaphora, or whatever), and yet decide that despite its outperforming all its rivals, it was not sufficiently accurate to warrant "the tentative conclusion that the statements of the best theory of the language faculty are true" (Chomsky, 1986, 249).[12]

Thus, when evaluating a given (small bit of a) theory of FHL against various rivals, it doesn't suffice to justify an ordinal ranking that places the favored theory on top. In general, that theory must also be "good enough". What this latter notion amounts to is a highly variable and holistic, not the sort of thing that is readily tractable. But as the example from confidence intervals suggests, it may be possible to chip off some simpler elements, ideally ones that may be assessed with methods that can be re-deployed, in a fashion that permits standardization and comparisons. The more this can be done, the more the results will lead to increasingly specific questions, as was seen in (7). E.g., we might start with:

(8) In developing or defending a linguistic theory as "best", what source(s) of uncertainty were aggregated, and how? How does the resulting combined uncertainty (in each case, if there is more than one) bear on the relevant bits of the theory in question?

In the case of confidence intervals, the questions in (8) all have straightforward answers; cf. any introductory statistics textbook. Moreover, even this elementary example moves us from the very general matter of which theories are to be accepted, to an intermediate one of which ones are good enough, to some even narrower ones of how various sources of uncertainty were combined, and how the composite was measured (presumably in a way comparable with its competitors). These latter issues are only tiny fragments of the original question about theory confirmation. But their diminuitiveness also suggests their explicit tractability, which we saw above is key to developing, analyzing, criticizing, and improving methods.

I conclude this section with two comments. First, with respect to the example just given, some researchers balk at the use of classical confidence intervals as a measure of uncertainty. Bayesians and others complain that they should be replaced with better measures, like posterior highest density regions. But from the present perspective, this only reinforces the basic position about needing greater explicitness. Such charges can be levied against classical intervals, precisely because they are fully explicit. It will be a true sign of advancement when corresponding means of assessment in linguistics are made so explicit that they too are subject to similar criticism.

---

[10] In general, the methodological challenge for linguistics is to improve its theoretical grip on *some* of the central sources of uncertainty in linguistic theorizing. Obviously, a theory can be undermined from practically any direction, so the in-principle sources of uncertainty are, in total, as global as the possibilities of Cartesian skepticism. But some sources of uncertainty tend, as a matter of fact, to appear more frequently than others, which is typical. E.g., in your standard psychology experiment, there is both (i) uncertainty from using a relatively small sample of subjects, and (ii) uncertainty from sampling without replacement from a finite population. (i) is the sampling error that structures the details of many experimental/observational designs; (ii), while also potentially serious, is often ignored by assuming that the sampling is genuinely random – i.e., that it is appropriately represented by a collection of i.i.d. random variables.

[11] Along these lines, Lee (Lee, 2004, 121) quotes Schlaifer:

> A point about hypothesis tests well worth making is that they 'are traditionally used as a method for testing between two terminal acts [but that] in *actual practice* [they] are far more commonly used [when we are] *given* the outcome of a sample [to decide whether] any final or terminal decision [should] be reached or [whether judgement should] be suspended until more sample evidence is available'.

[12] There is no need to quibble here over the notion of explanatoriness in the claim that we can adopt a realist attitude towards "constituent elements of the states postulated in an *explanatory* theory of behavior and enter into our best account of this behavior" (Chomsky, 1986, 253, emphasis added).

Second, any serious attempt to address such questions about evidence and methods (cf. (7) and (8)) will be heavily quantitatively based. There is, to my knowledge, no serious alternative. Indeed, Chomsky seems to support such a trend towards quantitative understanding of linguistic theorizing with such recent remarks as "*How much* of language [the operation of Merge] accounts for we don't really know – basically, finding that out is the Minimalist Program: *how much* is accounted for by this one innovation?" (Chomsky and McGilvray, 2012, 42, emphases added). I think this question is an excellent one. But the only hope for a serious answer to it will be quantitative.[13]

Such trends towards quantitative methods are only to be expected. The great statistician John Tukey, in an address to psychologists, summed up these matters well:

> The physical sciences have learned much by storing up amounts, not just directions. If, for example, elasticity had been confined to 'When you pull on it, it gets longer!' Hooke's law, the elastic limit, plasticity, and many other important topics could not have appeared. . . . The qualitative properties of *things* have proved much less important than the quantitative ones. Why should this not hold true for people? I believe that just this will prove to be so, but not without much effort. Bear in mind a simple fact: the great majority of the useful facts that physics has learned and recorded in numbers are specific and detailed, not global and general.... To move in this direction.... requires us to emphasize learning the right things to measure, learning empirically and by small painful steps.
>
> (Tukey, 1969, 86)

Tukey was also keenly aware that the difficult process of accomplishing this was far from complete. Elsewhere he stressed that the assistance supplied by imperfect and only partially explicit methods can often nevertheless provide a type of guidance that unaided human beings are unable to supply. This matter is addressed in the following section.

6. A meso-level of analysis

Above (Section 3), we saw that the geometric interpretation of PCA facilitated its subsumption and generalization under a more general geometric technique. This illustrates a key idea behind meso-level perspectives: they can help uncover useful similarities amongst various methods. In this section, I suggest that linguistics can gain much from seeking out potentially methodologically related fields and – with the same greed and hunger ascribed to psychologists in the quote that begins this paper – drawing hints and suggestions from them. In some sense, the example above of confidence intervals illustrates this. The present section expands on this with two further general points.

The first point is simply to emphasize just how very strongly the previous section's main claims are supported by the specific ways that other disciplines ply their trades. A look beyond linguistics shows that it is simply routine to explicitly attend to such features of the empirical data as its overall size.[14] Moreover, it is common for methodologically oriented books, journals, graduate courses, personnel, etc. to focus almost exclusively on fine-grained, explicit, and typically quantitative details; cf. the examples of journals in Section 1. Your typical discussions of formal methods in linguistics simply don't address the issues mentioned above (e.g., Johnson, 2008; Partee et al., 1990; Schütze, 1996).

Importantly, this basic point, which was detailed in Section 5, is largely independent of much of the currently popular "biolinguistic" program in linguistics. Generally speaking, the latter approach seeks to view FHL from a biological perspective, perhaps even substantially explained by general biological (or chemical, physical, etc.) factors not specific to language (Johnson, forthcoming; Lenneberg, 1967; Chomsky, 1995, 2005; Hauser et al., 2002). Although the nature of a theory can impact on the available means for studying it (and vice versa), many of the advances in biolinguistics should not be confused as *additionally* doing kinds of methodological work discussed here.

The challenges for linguistics are not so novel as to be unprecedented. For example, the economist Robert Sugden noticed a problem similar to the issue of redundancy (Section 5) when one tries to enumerate the "opportunities" one has at a given point in time. E.g., if you have the choice between beer, water and wine, it would seem you have more opportunities than if you have the choice between four different cans of Heineken beer, distinguished only by their serial codes (Sugden, 1998, pp. 317–318). In a more general vein, Milton Friedman, the 1976 Nobel laureate in economics, described in his famous 1953 essay, "The Methodology of Positive Economics" the methodological situation in economics as it was then, and where it should try to go.[15] Friedman paints a strikingly similar picture to that of linguistics, right down to issues of a theory "guiding" or merely "fitting" economic reality (Friedman, 1953, 157–8). There are probably several reasons for certain similarities between linguistics and economics, not least of which is that both fields rely heavily on large amounts of highly complex observational data that is difficult to measure precisely. It's beyond dispute that economic methods have grown increasingly explicit, which has made them easier to study, criticize, and improve. With a little luck, linguistics may soon follow a similar path. Moreover, if the extremely explicit algorithmic hypotheses about FHL are on the right track, one might even imagine a future where

[13] The reasons for quantification are very deep and general ones of measurement theory. Suffice it to say that they apply extremely broadly, to issues of length, mass, wealth, evidence, etc. E.g. "A unit was a fixed amount of a variable that could be repeated without modification and counted. The invention of units allowed the question *how much?* to be answered by counting *how many* units" (Masters, 2005, 18). (Crucially, of course, this latter question does not require integer- or rational-valued answers; hence the invention of the real numbers.)

[14] Indeed, a look *within* certain areas of linguistics beyond those discussed in this paper (phonetics, psycholinguistics, corpus linguistics, etc.) would support the same conclusions.

[15] A leading philosopher of economics describes this essay as "the most influential work on economic methodology of this century" (Hausman, 2008, 145).

highly complex fields like economics, sociology, biological systems research, etc., look to linguistics for guidance and suggestions.

I turn now to the second point of this section, which concerns the basic theory of FHL, sketched in Section 2. Importantly, FHL, I-languages, I-expressions and the like are unobserved. We only directly encounter combinations of I-language expressions along with any number of confounding influences, and various judgments about these combinations. Viewing linguistics in this way immediately suggests a potentially useful methodological point of comparison: latent variable models.

Like linguistics, a typical goal of using latent variable models is to extract complex interactive unobserved structure from complex interactive observed data sets that are themselves only indirectly related to the former. The earliest forms of latent variable models were first developed over a century ago, particularly for theories of intelligence (Pearson, 1901; Spearman, 1904; Bartholomew, 2004). Since then, a very broad range of other fields – including but not limited to economics, sociology, medicine, biology, chemistry, atmospheric and oceanic science, industrial engineering – have co-opted and modified them to suit specific needs (Bartholomew and Knott, 1999; Basilevsky, 1994; Brereton, 2003; Bollen, 1989; Jolliffe, 2010; Gould, 1996; Malinowski, 2002). That is, researchers in these fields recognized enough similarities between their methodological needs and those of others that they were able to exploit them and capitalize on their (sometimes distant) neighbors' successes. These similarities were, of course, data analytic methodological ones; after all, there's not much shared *empirical* content between theories of personality, urinary compounds relevant to animal path-marking, geological formations, stock market prices, blood chemistry, sources of pollution in a river, etc. There is, it turns out, substantial *methodological* overlap in how these areas are studied. Such similarities represent a bevy of extremely clear cases of highly salubrious meso-level perspectives. I suggest that latent variable models may similarly contain hints for developing an explicit linguistic methodology.

The discrete nature of both the manifest data and the underlying latent structure can help refine an initial search for hints. In particular, these features match some of the underlying requirements for latent class analysis and the analysis of other discrete items like contingency tables (e.g., Lazarsfeld and Henry, 1968; Greenacre, 2007; Hagenaars and McCutcheon, 2002). That said, there's no suggestion that any such techniques can be simply grafted onto a linguistic project with no further effort or insight. Quite the opposite. Rather, such methods provide some initial straws for linguists to grasp at as problems and means for approximate solutions are hopefully brought into better correspondence.

It might be thought that latent variable models have little to offer linguistic theorizing. After all, the former techniques are statistical and thus involve inherently random components; I-languages, by contrast, are inherently deterministic, with no such randomness. In reply, at least three points can be made. First, as was emphasized at the outset (Section 2), the present issue concerns not any particular theory of FHL, but rather our means for assessing them. This latter task is riddled with uncertainty – e.g., the relevant empirical facts are only partly known, and they are not fully separated from confounding influences; there are punishingly few constraints on the range of potentially correct theories of FHL, etc. Latent variable models may help manage that uncertainty in the evaluation of rival deterministic models. Second, it's not uncommon to study deterministic systems using nondeterministic means. As an extreme example of this, the great mathematician Paul Erdoš and his followers have made great use of probabilistic methods to prove theorems about the natural numbers. Perhaps more to the point, highly complex deterministic systems, like neural nets, are often tractable only via simplifying models that represent some of their behavior stochastically. Thus, there is no a priori reason to suppose they couldn't help reign in other systems like I-languages. Finally, many "statistical" techniques are themselves fully algebraic, or can at least be interpreted without using random variables. E.g., PCA, the singular value decomposition, correspondence analysis and many other methods can represent (literally: re-present) data, often in a reduced form, so as to reveal its internal structure in a way that simple visual inspection cannot.

I end with a note of caution. At the end of their book on latent variable methods, Bartholomew and Knott consider some modeling methods of a much simpler nature than what that linguist faces. They conclude with the warning:

> in the model-based approach to statistical methodology, everything is conditional on the model being true. When we come to models for relationships between latent variables we have reached a point where so much has to be assumed that one might justly conclude that the limits of scientific usefulness have been reached if not exceeded.
>
> (Bartholomew and Knott, 1999, 190)

Obviously, these warnings apply to linguistics, which seeks to uncover a highly intricate system of latent interacting elements that involve or interface with FHL. Using raw data with some unusual features, the linguist's task is to reverse engineer back to the I-expressions they reflect, and then back to the I-language that produced the latter, and finally to FHL, the cognitive organ of ultimate interest. In this sense we might characterize the linguist as working with *hierarchical* models with three levels of latent structure: the I-expressions, I-languages, and FHL. But when made explicit, hierarchical latent structure models can easily be so complex as to be nearly impossible to work with. There are many reasons for this. One is that they can be enormously flexible, to the point where they can accommodate almost any data that appears. The fact that the model fits the data is then a mathematical artifact of its flexibility, not of its similarity to the empirical phenomena. In philosophical terms, the structure of the model fits the empirical facts, but there is no reason to think that it guides them, in the sense of supplying any causal/explanatory story.

I do not know whether the methods of linguistics run the risk of being so powerful and flexible that its theories will be unable to supply much evidence about the existence and nature of FHL. Barring greater explicitness, it's not obvious that this question is even well-defined. But the risk is certainly there, whether acknowledged or not. If linguistic methods are too powerful in this way, it doesn't mean the general Chomskyan program regarding FHL is wrong; it only means that there currently aren't methods for assessing it. But such challenges are part and parcel of dealing with explicit methods, whose limitations can be clarified. Moreover, even if corresponding difficulties with more implicit methods aren't as clearly exposed, this does not suggest that they don't exist.

7. Conclusion

The truism that details matter in empirical research extends to *how* theories are developed and investigated – and to the ways these methods themselves are assessed for their reliability, validity, etc. Although discussion of linguistic methods has tended to occur at the macro-level, much attention is needed at the more important micro-level. Currently, there are virtually no linguistic methods developed at this level of detail, as the foreignness of such elementary questions in (7) and (8) illustrates. Instead of continuing to rely on implicit, holistic methods of expert judgment, which is famously an unwise strategy, concrete steps can be taken. This paper has discussed three of them. Most generally, a quantitative theory of the uncertainty in linguistic theorizing is needed. Such a theory would be of great service in assessing the bearing of evidence on theory, including such specific matters as whether the "best" overall theory should be "tentatively accepted as true", or whether the evidence is simply too weak to support any such inference. No such explicit methods have a hope of getting off the ground without a great deal of attention to such mundane details as estimating the size of the data sets involved.

More such details and methodological issues can be uncovered by looking at neighboring fields with a Meso-level perspective. The latter can help here by exposing relevantly similar issues in other disciplines that are handled much more explicitly by micro-level methods. Of course, it is not sufficient to merely co-opt techniques that work well for other problems in other areas. Rather, the goal should be to understand the particular problems in linguistics well enough to motivate the methods for coping with them. But meso-level views can allow linguistics to peer over disciplinary walls, ones currently reinforced by a kind of quantitative/qualitative divide, for hints about how best to follow other fields and grow its own methods.

References

Arkes, H., 2003. The nonuse of psychological research at two federal agencies. Psychol. Sci. 14 (1), 1–6.
Baker, M., 2001. The Atoms of Language. Basic Books, New York.
Baker, M.C., 1988. Incorporation: A Theory of Grammatical Function Changing. University of Chicago Press, Chicago.
Bartholomew, D.J., 2004. Measuring Intelligence: Facts and Fallacies. Cambridge University Press, Cambridge.
Bartholomew, D.J., Knott, M., 1999. Latent Variable Models and Factor Analysis, second ed. Hodder Arnold, London.
Basilevsky, A., 1994. Statistical Factor Analysis and Related Methods. Wiley-Interscience, New York.
Bishop, M., Trout, J., 2002. 50 years of successful predictive modeling should be enough: lessons for philosophy of science. Philos. Sci. 69, S197–S208.
Bishop, M., Trout, J., 2005. Epistemology and the Psychology of Human Judgment. Oxford University Press, Oxford.
Boeckx, C., 2006. Linguistic Minimalism: Origins, Concepts, Methods, and Aims. Oxford University Press, Oxford.
Boeckx, C., Piattelli-Palmarini, M., 2005. Language as a natural object – linguistics as a natural science. Linguist. Rev. 22, 447–466.
Bollen, K., 1989. Structural Equations with Latent Variables. John Wiley and Sons, New York.
Brereton, R.G., 2003. Chemometrics: Data Analysis for the Laboratory and Chemical Plant. John Wiley and Sons, Inc., Hoboken, NJ.
Carey, S., 2009. The Origin of Concepts. Oxford University Press, Oxford.
Chomsky, N., 1976. Problems and mysteries in the study of human language. In: Kasher, A. (Ed.), Language in Focus: Foundations, Methods and Systems. D. Reidel Publishing Co., Dordrecht, Holland, pp. 281–357.
Chomsky, N., 1981. Lectures on Government and Binding. Foris, Dordrecht.
Chomsky, N., 1986. Knowledge of Language. Praeger, Westport, CT.
Chomsky, N., 1995. The Minimalist Program. MIT Press, Cambridge, MA.
Chomsky, N., 2000. New Horizons in the Study of Language and Mind. Cambridge University Press, Cambridge.
Chomsky, N., 2002. On Nature and Language. Cambridge University Press, Cambridge.
Chomsky, N., 2005. Three factors in language design. Linguist. Inq. 36 (1), 1–22.
Chomsky, N., 2007. Of minds and language. Biolinguistics 1, 9–27.
Chomsky, N., McGilvray, J. (Eds.), 2012. The Science of Language: Interviews with James McGilvray. Cambridge University Press, Cambridge.
Cohen, J., 1994. The earth is round ($p < .05$). Am. Psychol. 49 (12), 997–1003.
Cox, D.R., 1958. Some problems connected with statistical inference. Ann. Math. Stat. 29 (2), 357–372.
Culicover, P., Jackendoff, R., 2010. Quantitative methods alone are not enough: response to gibson and fedorenko. Trends Cogn. Sci. 14 (6), 234–235.
Danks, D., Eberhardt, F., 2009. Conceptual problems in statistics, testing and experimentation. In: Symons, J., Calvo, P. (Eds.), The Routledge Companion to Philosophy of Psychology. Routledge, New York, pp. 214–230.
Dawes, R., 1979. The robust beauty of improper linear models in decision making. Am. Psychol. 34, 571–582.
DiSciullo, A.M., Boeckx, C. (Eds.), 2011. The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty. Oxford University Press, Oxford.
Dummett, M. (Ed.), 1993. The Seas of Language. Clarendon Press, Oxford.
Dunteman, G.H., 1989. Principal Components Analysis. Sage Publications, Newbury Park, CA.
Eckart, C., Young, G., 1939. A principal axis transformation for non-hermitian matrices. Bull. Am. Math. Soc. 45, 118–121.
Epstein, S.D., 2007. Physiological linguistics, and some implications regarding disciplinary autonomy and unification. Mind Lang. 22 (1), 44–67.
Fabrigar, L., MacCullum, R., Wegener, D., Strahan, E., 1999. Evaluating the use of exploratory factor analysis in psychological research. Psychol. Meth. 4 (3), 272–299.
Fitch, W.T., 2010. The Evolution of Language. Cambridge University Press, Cambridge.
Friedman, M. (1953). The methodology of positive economics. In (Hausman, 2008), pp. 145–178.
George, A., 1989a. How not to become confused about linguistics. In: George, A. (Ed.), Reflections on Chomsky. Blackwell's, Oxford, pp. 90–110.
George, A. (Ed.), 1989b. Reflections on Chomsky. Blackwell's, Oxford.
Gibson, E., Fedorenko, E., 2010. Weak quantitative standards in linguistics research. Trends Cogn. Sci. 14 (6), 233–234.
Gibson, E., Fedorenko, E., 2013. The need for quantitative methods in syntax and semantics research. Lang. Cogn. Process. 28 (1/2), 88–124.
Glymour, C., 1980. Theory and Evidence. Princeton University Press, Princeton, NJ.

Gould, S.J., 1996. The Mismeasure of Man. W.W. Norton and Co., New York (revised and expanded ed.).
Greenacre, M., 2007. Correspondence Analysis in Practice, second ed. Chapman and Hall/CRC, Boca Raton, FL.
Grimshaw, J., 1990. Argument Structure. MIT Press, Cambridge, MA.
Gross, S., Culbertson, J., 2011. Revisited linguistic intuitions. Br. J. Philos. Sci. 62, 639–656.
Grove, W., Meehl, P., 1996. Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical–statistical controversy. Psychol. Public Policy Law 2, 293–323.
Hagenaars, J.A., McCutcheon, A.L. (Eds.), 2002. Applied Latent Class Analysis. Cambridge University Press, Cambridge.
Hale, K., Keyser, S.J. (1986). Some transitivity alternations in english. Lexicon Project Working Papers 7, MIT, Center for Cognitive Science.
Hale, K., Keyser, S.J. (1987). A view from the middle. Lexicon Project Working Papers 10, MIT, Center for Cognitive Science.
Hatfield, G., 2002. Psychology, philosophy, and cognitive science: reflections on the history and philosophy of experimental psychology. Mind Lang. 17, 207–232.
Hauser, M., Chomsky, N., Fitch, T., 2002. The faculty of language: what is it, who has it, and how did it evolve? Science 298, 1569–1579.
Hausman, D. (Ed.), 2008. The Philosophy of Economics: An Anthology, third ed. Cambridge University Press, Cambridge.
Higginbotham, J. (1989). Knowledge of reference. In (George, 1989b), pp. 153–174.
Horn, R.A., Johnson, C.R., 1985. Matrix Analysis. Cambridge University Press, Cambridge.
Horn, R.A., Johnson, C.R., 1990. Topics in Matrix Analysis. Cambridge University Press, Cambridge.
Hornstein, N., 2009. A Theory of Syntax: Minimal Operations and Universal Grammar. Cambridge University Press, Cambridge.
Howson, C., Urbach, P. (2006). Scientific Reasoning: The Bayesian Approach, third ed. Carus Publishing Co., Chicago, IL.
Jackendoff, R., 2002. Foundations of Language. MIT Press, Cambridge, MA.
Jackendoff, R., 2011. What is the human language faculty? two views. Language 87 (3), 586–624.
Johnson, K., 2007. The legacy of methodological dualism. Mind Lang. 22 (4), 366–401.
Johnson, K., 2008. Quantitative Methods in Linguistics. Blackwell's.
Johnson, K. (2009). The need for explicit inferential methods in linguistics. In Dreyer, C.R. (Ed.). Language and Linguistics Emerging Trends. Nova Science, New York. pp. 193–208.
Johnson, K., 2011. A lot of data. Philos. Sci. 78 (5), 788–799.
Johnson, K. (Forthcoming). Notational variants and invariance in linguistics. Mind Lang.
Jolliffe, I., 2010. Principal Component Analysis, second ed. Springer-Verlag, New York.
Kayne, R., 1994. The Antisymmetry of Syntax. MIT Press, Cambridge, MA.
Koehler, D.J., Harvey, N., 2007. Blackwell's Handbook of Judgment and Decision Making. Blackwell, Oxford.
Lazarsfeld, P.F., Henry, N.W., 1968. Latent Structure Analysis. Houghton Mifflin Company, Boston.
Lee, P.M., 2004. Bayesian Statistics, third ed. Hodder Arnold, New York.
Lenneberg, E., 1967. Biological Foundations of Language. John Wiley and Sons, Inc., New York.
Magnus, J.R., Neudecker, H., 2007. Matrix Differential Calculus With Applications in Statistics and Econometrics, third ed. John Wiley and Sons, New York.
Malinowski, E., 2002. Factor Analysis in Chemistry. John Wiley and Sons, New York.
Marr, D., 1982. Vision. Freeman, San Francisco.
Masters, G.N., 2005. Objective measurement. In: Alagumalai, S., Curtis, D.D., Hungi, N. (Eds.), Applied Rasch Measurement. Springer, Dordrecht, pp. 15–25.
Mayo, D., 1996. Error and the Growth of Experimental Knowledge. University of Chicago Press, Chicago.
Meehl, P., 1954. Clinical versus Statistical Prediction: A Theoretical Analysis and Review of the Evidence. University of Minnesota Press, Minneapolis, MN.
Meehl, P., 1986. Causes and effects of my disturbing little book. J. Pers. Assess. 50 (3), 370–375.
Partee, B.H., ter Meulen, A., Wall, R.E., 1990. Mathematical Methods in Linguistics. Springer, The Netherlands.
Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. Philos. Mag. 2, 559–572.
Phillips, C. (2009). Should we impeach armchair linguists? In: Iwasaki, S., Hoji, H., Clancy, P., Sohn, S.-O. (Eds.). Japanese/Korean Linguistics, vol. 17. CSLI Publications, Stanford, CA, p. xxx
Royall, R., 1997. Statistical Evidence: A Likelihood Paradigm. Chapman and Hall, Boca Raton, FL.
Schott, J.R., 2005. Matrix Analysis for Statistics, second ed. Wiley-Interscience, New York.
Schütze, C., 1996. The Empirical Base of Linguistics. University of Chicago Press, Chicago.
Spearman, C., 1904. 'General intelligence', objectively determined and measured. Am. J. Psychol. 15 (2), 201–292.
Sprouse, J., Almeida, D. (2010). A quantitative defense of linguistic methodology. <http://ling.auf.net/lingBuzz/001075>.
Sprouse, J., Almeida, D. (2012). Assessing the reliability of textbook data in syntax: Adger's Core Syntax. J. Linguist. 48, 609–652.
Sugden, R., 1998. The metric of opportunity. Econ. Philos. 14, 307–337.
Suppes, P., 1962. Models of data. In: Nagel, E., Suppes, P., Tarski, A. (Eds.), Logic, Methodology, and Philosophy of Science: Proceedings of the 1960 International Conference. Stanford University Press, Stanford.
Trout, J., 1998. Measuring the Intentional World: Realism, Naturalism, and Quantitative Methods in the Behavioral Sciences. Oxford University Press, Oxford.
Tukey, J.W., 1969. Analyzing data: sanctification or detective work? Am. Psychol. 24, 83–91.
Uttal, W.R., 2003. Psychomythics: Sources of Artifacts and Misconceptions in Scientific Psychology. Lawrence Erlbaum Associates, Mahwah, NJ.
v. O. Quine, W. (1972). Methodological reflections on current linguistic theory. In Davidson, D., Harman, G. (Eds.). Semantics of Natural Language. D. Reidel, Dordrecht, pp. 442–454.
Walenski, M., Ullman, M., 2005. The science of language. Linguist. Rev. 22, 327–346.
Weller, S.C., Romney, A.K., 1990. Metric Scaling: Correspondence Analysis. Sage Publications, Newbury Park, CA.
Wickens, T., 1995. The Geometry of Multivariate Statistics. Lawrence Erlbaum Associates, Mahwah, NJ.
Williams, E., 1980. Predication. Linguist. Inq. 11, 203–238.
Williams, E., 1983. Against small clauses. Linguist. Inq. 14, 287–308.
Wilson, M., 2006. Wandering Significance: An Essay on Conceptual Behavior. Oxford University Press.