In J. Myers (Ed.) In search of grammar: Empirical methods in linguistics. Language and Linguistics Monograph Series 48. Taipei, Taiwan: Language and Linguistics.

Methods in search of grammar, grammar in search of methods*

James Myers
National Chung Cheng University
2011/10/26

1. Introduction

Linguistics is the scientific study of language. Its core subject matter is mental grammar, that is, knowledge of linguistic structure. Mental grammar is shaped by universal human biology. The syntactic component of grammar is studied through native-speaker judgments of acceptability. The phonological component of grammar is studied through the analysis of lists of words and, occasionally, phrases.

At least these are the standard aphorisms in theoretical linguistics, explicitly stated or latent in mainstream practice. Eventually every budding linguist learns that all of these claims remain quite controversial, both inside linguistics and among researchers in the allied fields of cognitive psychology and neuroscience. Is grammar really mental, or just a convenient description of linguistic behavior? If there is a mental component, is it knowledge of linguistic structure ("knowledge that") or of language use ("knowledge how")? Is it in fact grammar that biology shapes, rather than other aspects of the language learning and processing system? Are native-speaker judgments consistent and unbiased enough to reveal reliable information about linguistic knowledge? Do patterns in word lists provide information about productive grammatical knowledge in contemporary speakers, or only about the history of the words themselves? Above all, can we honestly call linguistics scientific if there is so much controversy about such basic issues?

The confusion facing linguistic theory is reminiscent of the state of pre-Newtonian optics, vividly sketched by Kuhn (1970:12-13):

> Each of the corresponding schools derived strength from its relation to some particular metaphysic, and each emphasized, as paradigmatic observations, the particular cluster of optical phenomena that its own theory could do most to explain. Other observations were dealt with by ad hoc elaborations, or they remained as outstanding problems for further research.... Being able to take no common body of belief for granted, each writer on physical optics felt forced to build his field anew from its foundations. In doing so, his choice of supporting

* Work on this chapter was supported by National Science Council (Taiwan) grants NSC 95-2411-H-194-005, NSC 95-2411-H-194-002, NSC 97-2410-H-194-067-MY3. Thanks to Szu-wei Stacy Chen, Chien-Jer Charles Lin, Paul Saka, Chentsung Yang, and an anonymous reviewer for their comments. The usual caveats apply.

> observation and experiment was relatively free, for there was no standard set of methods or of phenomena that every optical writer felt forced to employ and explain.

For "optics" read "linguistics". Oversimplifying for rhetorical effect (as Kuhn presumably was doing as well), here we find one prominent school that takes native speaker judgments as paradigmatic evidence, while other schools challenge this school (and each other) with their own favored data sources (language processing, corpus data). Each school is satisfied to build its own version of linguistics in relative isolation from all of the others, with no generally agreed-upon methods for choosing among them.

Rather than bemoaning this sorry situation yet again (for a recent example of hand-wringing over the state of modern linguistics, see Carlson 2003), this book will instead ask how linguists might go about getting out of it. The guiding idea here is that scientific consensus relies on the one thing that all scientists undeniably have in common: observable reality. Newton did not have a unifying effect on optics solely through the sheer force of his ideas (some of which were immediately questioned; Shapin 1996). Rather, what won Newton's contemporaries over to his optical theory was his empirical argument for it, wherein he demonstrated certain undeniable facts through a series of experiments - cleverly designed, elegantly analyzed, and above all, replicable.[1] The best way to encourage greater cohesiveness in linguistic theory, then, is to encourage linguists to feel some responsibility towards the whole wide range of linguistic data, not just their own favorite sorts.

Though theoretical linguists have never entirely neglected "non-traditional" data sources, only in the past couple of decades has the trend begun to enter the mainstream, turning the non-traditional into the traditions of the future. The movement in this direction seems to have started to gain momentum a little over twenty years ago, in phonology, when Ohala and Jaeger (1986) and Kingston and Beckman (1990) demonstrated the advantages of adopting experimental, corpus-based, and quantitative methods from psycholinguistics and phonetics to the testing of hypotheses in theoretical phonology. The syntacticians entered the game a few years later, when the publication of Schütze (1996) and Cowart (1997) sparked new interest in improving the collection and analysis of native-speaker judgments in syntax via stricter experimental standards adapted from cognitive psychology. Since then many anthologies (including journal special issues) have been published on empirical methods in theoretical linguistics, including Borsley (2005), Coetzee, Kager, and Pater (2009), Cole and Hualde (2007), Everaert, Lentz, De Mulder, Nilsen, and Zondervan (2010), Everaert and Musgrave (2009), Fanselow, Fery, Schlesewsky, and Vogel (2006), Featherston and

[1]Historians and philosophers of science will recognize this version of the story as rather simplistic. Among other things, scientists surely share not just observable reality, but also strong biases about what makes a theory convincing. Nevertheless, while scientists often disagree about theory and even methodology, replicable observation seems to be a universally agreed upon virtue.

Sternefeld (2007), Featherston and Winkler (2009), Gonzalez-Marquez, Mittleberg, Coulson, and Spivey (2007), Kepser and Reis (2005), Krifka (2007), Meibauer and Steinbach (2011), Penke and Rosenbach (2007a), Rice and Newman (2010), Ritter (2002), Solé, Beddor, and Ohala (2007), Stefanowitsch (2007), Stefanowitsch and Gries (2007), and Winkler and Featherston (2009). Even more importantly, perhaps, it is becoming ever more commonplace for experimental, corpus-based, and/or computational studies to appear in the core linguistics journals alongside the more "traditional" articles.

Yet this revolution in methodology has yet to have much effect on how linguistics is taught. The dedicated student can certainly learn a lot by carefully examining, and then attempting to emulate, the many recent research studies applying non-traditional data sources to grammatical questions, but this may not be the easiest way to start expanding one's methodological toolbox. After all, most of these studies are not intended to serve a pedagogical purpose, and many already assume a certain level of familiarity with experimental or quantitative methods. Some specialized tutorials do exist, including introductions to quantitative corpus linguistics (e.g. Kennedy 1998, Manning and Schütze 1999, Gries 2009) and Cowart's (1997) introduction to syntactic judgment experiments, but as yet there is no general resource providing gentle introductions to the wide range of empirical tools available for testing grammatical hypotheses.

This book provides a step in this direction. In one concise volume, it collects not only applications of non-traditional empirical methods in a variety of original linguistic studies, but also step-by-step tutorials, rich in examples, on how linguists may apply such methods in their own research. The contributions are selected to provide a balanced overview of the methodological landscape. Thus the book covers experimental, corpus-based, and computational approaches, addresses issues in syntax, morphology, and phonology, and includes data from English, Chinese, and a smattering of other languages.

Before introducing these contributions, in section 2 we first revisit the question of why linguists should worry about methods in the study of grammar. Section 3 asks what grammar is anyway, why it cannot be observed directly, and why the standard methods we use to study it have remained standard for so long. Finally in section 4, we take a peek into the future of linguistic methodology hinted at by the contributions in this book.

2. Methods in search of grammar

Arguing that linguists should improve their methods may seem as irrelevant as arguing that the sun should rise; of course linguists will improve their methods in the course of time, since that's just the nature of science. After all, nobody could deny that the range of methods available to linguists today is a vast improvement over those available one hundred years ago. Yet arguments over methodology are also part of science. Methodological disagreements are

particularly lively in social sciences like linguistics, presumably because human behavior is less well-behaved than the systems studied by physicists, chemists, and biologists.

Nevertheless, there are still many linguists who don't see what all the fuss is about. The traditional methods (traditional at least since the generative revolution of the 1950s) are working just fine, they say; supplementing or replacing them with more rigorous methods is more work than it's worth. Newmeyer (1983, 2007), Phillips and Lasnik (2003), and Phillips (2009) make this point explicitly with respect to informal acceptability judgments in syntax; Chomsky and Halle (1968) and Bromberger and Halle (1989) justify traditional phonological methods (i.e. studying word patterns in dictionaries) on philosophical grounds; Pullum (1983/1991) argues more generally that linguists shouldn't waste time on metascientific issues at the expense of doing actual linguistic research.

This section attempts to pop this bubble of complacency with a two-pronged fork.[2] First, in section 2.1, I examine the apparent source of the neglect of methodological issues in modern linguistics, namely the confusion between two distinct concepts: "empirical" and "empiricism". Then in section 2.2, I show how the lack of methodological rigor has repeatedly caused theoretical linguists into trouble.

2.1 A rational approach to empiricism

Calls to improve, or at least to think more deeply about, the methodology of linguistics have always been with us, but they tend to come with a certain amount of controversy. One possible reason for this is that they frequently get mixed up in an ancient debate over the source of human knowledge.

On the one side is empiricism, the view that knowledge derives primarily from our senses, advocated in contemporary linguistics most strongly by corpus linguists like Geoffrey Sampson (2001:1), who states that "if we want to deepen our understanding of language, our best way forward is to apply the same empirical techniques which have deepened our understanding of other observable aspects of the universe during the four centuries since Galileo." Sampson (p. 1) defines the "empirical scientific method" as a bottom-up process: "Listen, look. Summarize what you hear and see in hypotheses which are general enough to lead to predictions about future observations." For Sampson, this means that the only scientifically appropriate linguistic data are corpus data, given that "[l]anguage ... is a concrete, tangible aspect of human behaviour," that is, directly observable via the senses, unlike abstract, intangible native-speaker intuitions.

On the other side is rationalism, the view that knowledge derives from deductive reasoning rather than solely from experience, exemplified in contemporary linguistics by

[2]Featherson (2007) takes a similar tack, albeit using a different metaphor, urging empirical reform in syntax by offering both a carrot (the new theoretical insights made possible by improved methods) and a stick (the empirical mistakes that can arise when relying solely on traditional methods).

Noam Chomsky, who applies rationalism not only to babies but to scientists as well. To take just one recent quotation (see Botha 1989 for others), Chomsky (2002:102) describes the development of modern science in terms of "the Galilean move towards discarding recalcitrant phenomena if you're achieving insights by doing so, the post-Newtonian concern for intelligibility of theories rather than of the world, and so on. That's all part of the methodology of science." Thus according to Chomsky, science became science precisely by becoming less dogmatic about the primacy of observation.

This view may seem peculiar at first; as Chalmers (1999) points out (in a highly accessible introduction to the philosophy of science), the empiricist view, in which tangible, real entities force their way unbidden into our attention, is the naive view most people have about science works. Problems with this naive view become clear, however, when one recognizes the logical impossibility of defining "evidence" completely independently of "theory." Far from facts speaking for themselves, the recognition of evidence as real and relevant is a complex psychological and social process (as Kuhn and others have documented). Even raw observations require some interpretation; in the late nineteenth century Giovanni Schiaparelli observed channels on Mars, but they later proved to be optical illusions (Yule 1978). Another problem is that history's greatest scientists, from Galileo on down, have tended to be a bit more stubborn in their theoretical positions than a strictly bottom-up, ugly-facts-kill-beautiful-theory empiricist view would lead us to expect, and this stubbornness is at least partly responsible for their greatest achievements. Still another problem is that the link between observation and hypothesis always involves subsidiary hypotheses, which require their own observations for support, which then require further subsidiary hypotheses, and so on (this is sometimes called the Duhem-Quine thesis; Gillies 1998). Thus even data in corpus linguistics do not speak for themselves; interpreting them requires, among other things, subsidiary beliefs about the representativeness of corpus data and the link between language production and mental knowledge.

The most productive way out of the empiricist/rationalist debate is not to take sides but to try to get both sides to work together. Note the peculiar fact that Sampson invokes Galileo as the paradigmatic objective observer, while Chomsky praises him for playing fast and loose with facts. Historically, both views about Galileo are correct (see Botha 1982), and they are true more generally as well: modern science is neither pure rationalism nor pure empiricism, but a novel (if sometimes awkward) synthesis of these two ancient traditions. It's hard to maintain the balance, and linguistics is hardly alone in having scholars who tilt more towards one side than the other. Shapin (1996) contrasts the mathematically elegant Newton (another of Chomsky's heroes) with his contemporary Boyle, who was a far more careful experimentalist, yet so cautious about theorizing that he never even formalized the gas law that today bears his name. Shapin calls their contrasting approaches simply "different games that natural philosophers might wish to play" (p. 117).

Thus in advocating that linguists pay more attention to evidence, I emphatically do not favor empiricism over rationalism: one without the other just isn't science. Instead, my goals are purely pragmatic: observations are public, concepts are private, so consensus in science ultimately depends just a bit more on the former than the latter. Consensus in science thus requires some sort of general agreement about methodology, and just a bit more respect to the data favored by competing schools.

2.2 What could possibly go wrong?

In a defense of traditional syntactic acceptability judgments, Phillips and Wagers (2007:740) are careful to limit their discussion to the "best practices" in the discipline. Yet this leaves as open questions how well linguists keep up these high standards in general, and whether the risk of empirical blunders could be reduced by adopting more stringent methodological protocols. Phillips and Wagers (2007:740) themselves go on to note that "[i]t is not difficult to find instances of careless misrepresentation of linguistic data." In this section I cite several examples.

Starting with syntax, consider the Chinese sentence in (1). Huang (1982) claims that for syntactic reasons, *shenme* ("what") can have wide scope here, so that the sentence can mean "What was bought, such that you wonder who bought it?" Xu (1990) denies that this meaning is possible for this sentence. Who is right? How could we be sure?

(1) Ni xiang zhidao shei mai-le shenme?
you want know who buy-ASPECT what

Such disagreements over acceptability (see Schütze 1996, Featherston 2007, and Myers 2009a,b for further examples) may arise due to different idiolects (though Labov 1975, 1996 argues that the empirical evidence for idiolects, divorced from speech communities, is very weak), differences in speakers' thresholds for "acceptable" versus "unacceptable" (see Sprouse 2009 for data relating to this notion), differences in the theoretical interpretation of clear judgments (as argued in Newmeyer 1983, 2007), differences in sensitivity to imagined pragmatic context (Carroll, Bever, and Pollack 1981 even show how judgments can be affected by the presence vs. absence of a mirror, supposedly influencing how speakers see themselves as social beings), or random variations across speakers within a speech community, or indeed from one judgment to the next by the same speaker (a central issue in Cowart 1997).

The fact that judgment data and their interpretation can be so ambiguous has led some critics to dismiss them entirely (e.g. Sampson 2001, 2007). A more pragmatic approach, however, is to recognize their enormous value in generating hypotheses, and even in

resolving empirical debates when they are not ambiguous, while still working to understand and reduce their limitations. As argued by Schütze (1996, 2011), Cowart (1997), Featherston (2007), and Myers (2009a,b), this can be done by collecting and analyzing judgments in accordance with protocols building on the almost two centuries of experience that experimental psychologists have developed for dealing with their similarly variable and multifaceted behavioral response measures.

A key part of these protocols is experimental design, a point emphasized in Cowart (1997) and Myers (2009b). To illustrate what can go wrong when linguists fail to design informal judgment experiments carefully, consider the English judgment contrast in (2) claimed by Di Sciullo and Williams (1987:33). These examples are intended to show that affixes like *-er* are transparent to the assignment of semantic roles (*bread* is still the patient of *bake*), whereas roots like *man* are not.

(2) a. a baker of bread
b. *a bake-man of bread

However, the claim actually involves a relationship between two elements, *bake* and *bread*, and therefore the implicit design of the judgment experiment should involve two factors, not just *-er* versus *man*, but also *bread* versus no *bread*. Adding the [±*bread*] factor results in the two additional forms in (3), where (3b) is already worse than (3a), even without the presence of a patient argument. Moreover, the unacceptability of (3b) is no accident; as Spencer (1991:333) points out in his commentary on this failed argument, *man* generally doesn't attach to verbs, especially not transitive ones (see Myers 2007 for further discussion).

(3) a. a baker
b. *a bake-man

Phonologists also sometimes employ native speaker judgments, as when dealing with phrasal patterns. Thus a whole stratum in Halle and Mohanan's (1985) model of English lexical phonology is based on a rule of /l/ resyllabification which they justify using the phonetic intuitions given in (4) (based on their figure (20), p. 65).

(4) a. palatalized /l/: a whale edition
the seal office
b. velarized /l/: the whale and the shark
the seal offered a doughnut

Unfortunately for their analysis, Sproat and Fujimura (1993) showed, using instrumental

phonetics, that /l/ resyllabification actually seems to involve not a binary palatalized/velarized /l/ contrast but a complex interplay between prosody and the two subgestures involved in forming the lateral (lowering the tongue body and raising the tongue blade). Unsurprisingly, then, Hayes (2000) found that there was a considerable degree of variability and gradience in native-speaker acceptability judgments for palatalized versus velarized /l/. Given that these later studies were based on a larger amount of data, collected in a more systematic fashion that those cited by Halle and Mohanan, and then analyzed quantitatively, they are inherently more convincing.

The dominant traditional method in phonology involves not judgments but the analysis of word sets, taken from a dictionary or elicited from a native speaker. Like acceptability judgments in syntax, dictionary data have earned their central place in phonology by providing systematic arguments for a whole host of fundamental theoretical concepts and are unlikely to disappear from the phonologist's toolbox any time soon. Yet detecting patterns in a data set is essentially a statistical problem, and phonologists who neglect to quantify can get stuck.

A striking example of this is the confusion over tone spreading in Mende. On one side, Gussenhoven and Jacobs (2011:159) want to show that in this language, tone consistently spreads left to right, and so they choose examples from Leben (1978) like those in (5), with surface high-low-low and low-high-high patterns. By contrast, Zoll (2003:230-1) wants to show that the direction of tone spread in Mende depends on tone quality (high vs. low), and so she chooses examples from Leben (1978) like those in (6), with surface high-low-low and low-low-high patterns. Presumably the examples chosen by these authors are all genuine Mende words, but only Gussenhoven and Jacobs (2011) claim that the pattern in their sample represents "the majority of Mende words" (p. 159). This essentially statistical point seems quite relevant if we want to know which of these two analyses may be a more accurate description of Mende mental grammar.

(5) a. [félàmà] "junction"
b. [ndàvúlá] "sling"

(6) a. [félàmà] "junction"
b. [lèlèmá] "mantis"

Linguists are only human. Nevertheless, it seems likely that the risk of such problems could be reduced by paying more attention to methodology, both in the training of linguists and in the evaluation of their research by their peers.

3. Grammar in search of methods

Advocating better methods makes obvious sense if we are searching for generally recognized entities that merely happen to be hard to reach, such as underground oil reserves. Yet not only is grammar hidden, but its very nature is subject to much controversy, with some doubting that it even exists.

Not quite knowing what one is looking for is hardly a unique situation in science. As I point out in 3.1, history suggests that the most promising strategy is to swallow one's doubts and start somewhere anyway. Surprisingly perhaps, in the case of grammar, linguists of competing schools already tend to agree about several of its key features. One of these features, as discussed in 3.2, is that grammar can be inferred only indirectly via observed linguistic behavior. The methodological consequence of this is that we must proceed cautiously when building chains of inferences from observation to hypothesized grammar.

3.1 Where to start the search for grammar

It might seem natural to start the search for grammar by defining explicitly what we mean by grammar. However, practicing scientists rarely, if ever, begin this way. Instead, scientists generally start out with a vague notion of what they're looking for, and then as they work, they gradually shift their focus until they may end up somewhere quite different from where they were first headed.

Churchland (2002), who is interested in the study of consciousness, a concept even more elusive than grammar, illustrates this point with a number of historical examples. Take "fire" for instance. Initially (and still naively today), this concept seemed to apply both to a candle and to the sun, since both give off heat and light. Today we know, however, that the sun's heat and light are the result of nuclear reactions whereas the candle's result from the chemical reaction of oxidation. Thus the candle's flame actually classes with rust, also caused by oxidation, rather than with the sun. If scientists had insisted all along that, by definition, the candle and the sun must be identical, they may never have discovered that they actually weren't. In the same way, we don't know exactly what "grammar" will turn out to be in reality, and so it would be a bad idea to fix a definition for it before starting the search.

But how can it be possible to study something without defining what it is that we are studying? Again, history can be our guide. The notion of fire began with relatively intuitive and pre-theoretic observations about the world. By analogy, then, our quest for grammar should start with views about it that are arguably most "intuitive" (at least among linguists).

Despite the rancor across the various schools of linguistics, it is surprisingly easy to come up with a list of features that most linguists would agree to hold of "grammar." First, grammar is that which is covered in a traditional descriptive grammar book, namely pronunciation (phonology) and the formation of words and sentences (morphosyntax).

Grammatical analyses have existed for centuries and in many cultures, from Babylonian descriptions of Sumerian (Civil 1994) to Pāṇini's analysis of Sanskrit (Cardona 1994) to traditional Chinese phonology (Downer 1963).

Grammar is also seen (by definition perhaps) as systematic (i.e. regular, lawful, productive, rule-governed, constrained, structured), which means that in principle it goes beyond rote memorization and ad hoc analogy (though the borderline between grammar and analogy is sometimes seen as fuzzy; see e.g. Albright and Hayes 2003). Seuren (1998) traces the notion of grammar as system surprisingly far back in the linguistic tradition in the West.

Another central aspect of the notion of grammar is that it is mental. This has been the explicitly stated mainstream view since the cognitive revolution of the 1950s, but its roots are older. Sapir (1933/1949) made the phrase "psychological reality" famous, Bloomfield (1914) assumed linguistics was a "mental science" (though in later work like Bloomfield 1933 he made a sharp division between the work done by linguists and by psychologists), and so on back through the centuries (again, see Seuren 1998). Long before the scholars got involved, folk psychology already saw a language as something that people "know." Mentalism is also embraced by the large contingent of Chomsky's detractors calling themselves cognitive linguists.

Some do still argue that the descriptive grammars of linguists aren't mentally real, though this view is more common among philosophers (e.g. Devitt 2006, Millikan 1995) and psychologists (e.g. Edelman and Christiansen 2003) than among linguists. Moreover, on closer examination, their argument is not so much that mental grammar doesn't exist at all, but more that linguists misdescribe it. Devitt (2006), for example, thinks that linguists' grammars are good descriptions of linguistic behavior; he merely challenges the notion of mental grammars as composed of the sorts of rules and principles of the descriptive grammars. Sampson (2007) goes a step further, challenging the adequacy of linguists' grammars for describing corpus data, but he still assumes some sort of mental machinery giving rise to the regularities in the corpus data. Derwing and Baker (1978) adopt a processing-oriented approach, advocating the replacement of Chomsky's (1965) notion of "competence" with a "competence to perform": that is, knowing language is a form of "knowing-how" (procedural knowledge), not "knowing-that" (propositional knowledge).

There is thus considerable agreement, across a variety of schools of thought, that there is indeed something real in the structure and/or processing of the brain that is correlated with the regularities observed in linguistic behavior. This something is what most linguists would call mental grammar, regardless of disagreements over its precise nature and how it is best formalized or studied. Interestingly, the know-that and know-how contrast is reflected in folk terminology. While English speakers say "I know English", in Chinese and other languages the idiom is literally "I can English" (我會英文), describing an ability rather than possession of information. Either way, language is universally recognized as inherently psychological.

Mental grammar is also universally understood to reside in permanent memory rather than being a temporary or rapidly evolving process (after initial acquisition). One practical application of this notion is that permanence can be used as a diagnostic to distinguish grammar from transient processes or representations that emerge while language is being used. This diagnostic is valid even if grammar is stored as a memorized set of procedures or habits, rather than as propositional rules; the short-lived processes involved in the parsing of some specific sentence are still distinguishable from the permanent knowledge of parsing strategies in general.

Moreover, despite a long history of disagreements over what is innate in grammar, the grammar of any particular language is nevertheless universally agreed to be partly learned. This property of particular grammars is sometimes employed to argue that some feature of linguistic behavior is governed by grammar, rather than merely being due to physics or general cognition. For example, we cannot say (as early linguists sometimes speculated) that the order of words in a sentence merely follows the natural order of our thoughts, or that the combination of sounds within words merely obeys the laws of physics, since languages can vary considerably in both properties, despite their speakers apparently having similar thoughts and speech organs (cross-linguistic variation in signed languages leads to similar conclusions).

Finally, and most importantly for a book on linguistic methodology, mental grammar is universally seen as hidden from direct observation. This view is implicit in the competence/performance distinction introduced in Chomsky (1965), according to which mental grammar (competence) is only available to observation via the output of language processing (performance). An implication of this, also introduced in Chomsky (1965), is that we cannot equate a native speaker's judgment of the (relative) acceptability of a sentence with the evaluation of this sentence in terms of (relative) grammaticality, since sentences can sound better or worse for reasons other than their grammatical status (e.g. they may be relatively harder or easier to process). Even critics of the notion of mental grammar agree with something like this distinction, at least the claim that it is linguistic behavior, not its mental underpinnings, that is observable. It's just that, for them, the unobservability of grammar implies that it is at best unknowable, a claim I will return to in the next section.

There are of course many disagreements about the nature of grammar, even beyond those noted above. Surely it is useful, however, to appreciate how much linguists already agree on, since this agreement is the foundation upon which a more comprehensive consensus will be built.

3.2 Forging the links between observation and grammar

Grammar is thus a (partly) learned, stable, mental system underlying observed linguistic

regularities. The tricky bit is getting at this system from these observations. The difficulty of justifying an inferential chain from overt data to hidden grammar has been both overestimated and underestimated, but there does seem to be an obvious strategy for dealing with it: focus at least part of one's research time to the inferential chain itself. One simple corollary of this (often neglected) truism is that different types of grammatical hypotheses may require different types of inferential chains, including different data sources.

Those who (claim to) reject any notion of mental grammar (even as the competence to perform) do so out of pessimism that such inferential chains can be justified at all. After all, if neither grammar nor the processes assumed to link it to the observable world are themselves observable, how can the chain be tested? This essentially behaviorist position has mostly faded away, since it is incompatible not just with mainstream linguistic theory but with the rest of contemporary cognitive science as well (though some advocates still exist; e.g. Schlinger 2004, Uttal 2007). One problem with this pessimistic position is that it undermines most, if not all, of science more generally. Not only can many entities in science not be observed with the naked eye, but hidden entities can come to be accepted as real even if there is nothing like them in direct experience (e.g. subatomic particles or dinosaurs). Chomsky (1980) cites the example of the interior of the sun, which we can study only indirectly from observation of the sun's surface. The very notion of direct observation is an illusion anyway, given all we know now (ironically, with the help of cognitive psychology) about how much computation is involved in even the simplest of perceptual experiences.

Nevertheless, there is a core insight in this criticism that is worth preserving, namely that a claim about an observation is ultimately only as convincing as the chain of reasoning linking the observation to the claim. Consider again Schiaparelli's Martian channels, where the weak link in the inferential chain was the blurriness of the key telescopic images; improved telescopes brought sharper images, and the channels disappeared. The same holds for early twentieth century skepticism over Alfred Wegener's theory of continental drift, which he based on the only evidence then available to him (e.g. matching fossil strata in Africa and South America). Even though this theory was eventually accepted, the skeptics had every right to be skeptical at first: the argument literally had a gaping hole in it, namely the Atlantic Ocean, without any testable mechanism for its creation. The inferential chain was not completed until evidence for the growth of the Mid-Atlantic Ridge became available in the second half of the twentieth century (see Cohen 1985 for details; this story also provides another example of how concepts shift in the maturation of a science, as continental drift became plate tectonics). As was noted in section 2.1, it is common for self-confident scientists to press on with claims even when intervening steps in their argument are missing, but the same can be said of self-confident cranks. Ultimately the gaps do have to be filled in.

The gaps have been filled in for, say, inferences from the surface of the sun to its interior (via the laws of chemistry and physics), with each step in the argument backed up by

laboratory experiments and other observations. We can hardly say the same about the inferential gaps between, say, acceptability judgments and underlying mental grammar. Nobody really knows how grammar causes acceptability sensations, or even whether the relationship is truly a causal one. There has been a lot of philosophical discussion about the problem (e.g. Devitt 2006, Schütze 1996, Matthews 1991, Neeleman and Van de Koot 2010), but surprisingly little empirical research (though see Bornkessel-Schlesewsky and Schlesewsky 2007, Luka and Barsalou 2005, Nagata 1989, 2003, Sprouse 2007, 2008, 2009, 2011, for some attempts to understand the nature of syntactic judgment making).

I would argue that the failure to recognize the challenge of the data-theory gap in grammatical theorizing deserves much of the blame for the gaffes cited in section 2.2. Though the view of grammar as only indirectly observable is mainstream dogma (Phillips and Lasnik 2003:61 call it a "truism"), it is often neglected in practice; there is a strong temptation to treat observed linguistic behavior as identical to grammar itself. One sign that linguists do in fact give into this temptation (as noted by Newmeyer 1983, 2007 and Schütze 1996, among others) is the ambiguous use of diacritics on linguistic examples, which sometimes represent acceptability (observed data) and sometimes grammaticality (claimed theoretical status).

To take an example essentially at random, Aronoff and Cho (2001:167) call the contrast in (7) an "observation", and their use of ?? rather than * implies that they believe that the observed acceptability can be measured to some degree of precision.

(7) a. penmanship [from (1), p. 167]
b. ??womanship

On the very next page, however, they claim that (8b) "is not well-formed" and describe a similar pair as a "grammatical contrast"; their use of a * versus unmarked contrast is consistent with an intention to represent a categorical grammatical claim rather than an observed difference in degree of acceptability.

(8) a. I hate it when my friends are older than me. [(2), p. 168]
b. *I hate it when my parents are older than me.

Another symptom of the temptation to equate grammar with observed linguistic data is the standard practice among phonologists to assume that the phonological component of mental grammar is responsible for describing all patterns attested in the lexicon. Yet everyone recognizes that lexicons have hundreds of thousands of arbitrary, extra-grammatical properties, from the existence of words for certain concepts and not others, to the specific sequences of segments and syllables that distinguish morphemes. This raises the possibility

that all lexical items are memorized as-is, obviating the need for synchronic mental grammar (Kenstowicz and Kisseberth 1979; Ohala 1986). Some would reject grammar diachronically as well, ascribing phonological patterns to the effect of gradually accumulated extra-grammatical misperceptions during acquisition (Blevins 2004; Hansson 2008). Recognizing these possibilities is not the same as accepting them, but ignoring them is naive: the line from phonological data set to mental grammar is no more direct than that from syntactic judgments.

One way to put all of this is to say (along with previous authors like Penke and Rosenbach 2007b) that there is no such thing as "competence data". All linguistic data are "performance data", simultaneously giving information both about systematic knowledge and about extinct or temporary processes. Thus every study on grammar is a psycholinguistic study, and should be as carefully planned and carried out. Experimental psycholinguists follow the same principle in reverse, controlling for "mere" grammatical status so that fleeting processes can be observed without interference from grammar (e.g. making sure that all sentences in a parsing experiment are equally well-formed, varying only in parsability). The raw data in a grammar study (say, testing a syntactic hypothesis) and a psycholinguistic study (say, testing a hypothesis about parsing) may thus be of exactly the same type (acceptability judgments or reading times), and many of the stimuli may be the same as well; it is only the research goals, and hence the experimental designs, that are different.

More commonly, of course, psycholinguists and grammarians employ different types of data and inferential chains. This is not entirely due to historical convention, but follows in part from differences in the types of hypotheses they consider. In particular, psycholinguists are interested in real-time processing, while grammar is assumed to be stable over long stretches of time. Hence it makes sense that the former would rely heavily on experiments measuring reaction times, while the latter would rely on off-line measurements like acceptability judgments.

The same sort of considerations are responsible, in part, for differences in methodological conventions across subdisciplines of grammar research. Why have syntacticians (at least since Chomsky 1957) relied more on acceptability judgments than corpus analyses? The answer has less to do with rationalism versus empiricism than some have implied (e.g. Sampson 2001), and more to do with combinatorics. Syntactic hypotheses are difficult to test in corpora because syntax is too productive. The typical language has thousands of vocabulary items and many different grammatical constructions, each typically several words long. The result is that in order to sample every one of these constructions with enough variety in vocabulary to show that the constructions are not word-specific, one would need a corpus of an unrealistically large size. Thus it is virtually guaranteed that there will be theoretically important grammatical structures that are so rare that a corpus analysis alone will have difficulty establishing that they are indeed grammatical, and not some sort of fluke.

By contrast, as we have seen, the analysis of dictionaries (databases compiled from corpora) is the dominant methodology in phonology. While the experimental approach is often claimed by linguists, of all stripes, to be inherently superior to mere observation (e.g. Chomsky 1957, Ohala 1986), non-experimental observation is an entirely respectable tool in such uncontroversially scientific fields as astronomy and paleontology (Cleland 2002). Moreover, the preference for this approach in phonology is not arbitrary. First, phonology does not face as big of a combinatorial challenge as syntax, since there are many fewer basic units (dozens of phonemes instead of thousands of words), and since (as an apparently empirical fact) phonological patterns tend to be more local, meaning that the analysis of shorter strings (i.e. parts of words) may suffice for their testing. Second, as a sort of corollary of the reduction in combinatorial possibilities, acceptability judgments are of less use in phonology than in syntax. There is clearly little point in asking people to judge the acceptability of existing words, since they may respond solely on the basis of rote memory, without engaging grammar. Yet judging nonwords doesn't provide particularly clear evidence for grammar either, since there is abundant evidence that people given this task also rely on superficial analogy with memorized words (e.g. Bailey and Hahn 2001). The analogy problem is not as serious in syntax since combinatorics makes it trivial to invent sentences that differ as much as one likes, in lexical content or overall structure, from anything that has ever been uttered before.

One sign that combinatorics is the primary reason for the different methodological traditions in syntax versus phonology is that morphology, which stands somewhere between sentence structure and phonological structure in number of basic units and combinatorial freedom, is typically studied using both methodologies. This is illustrated by the Aronoff and Cho (2001) paper mentioned above, which supplements the cited judgment data with data from a dictionary.

While these traditional methods are thus better justified than their critics sometimes give them credit for, they can still be improved upon. One straightforward way forward is simply to make them more precise, via proper statistical sampling methods and quantification. For example, there is now a large and growing literature on the collection and analysis of syntactic acceptability judgments using standard psycholinguistic techniques: sentence stimuli are presented in carefully controlled sets, in random order, to naive judges for their responses, which are then analyzed statistically, in an approach sometimes called experimental syntax (Cowart 1997; Featherston 2007; Myers 2009a,b; Schütze 2011). Similarly, traditional phonological analyses of dictionary data can be sharpened by counting the number of items that obey or violate hypothesized constraints (Uffmann 2006; Zuraw 2007; Myers 2009c).

Quantification alone will not deal with all of the limitations of the traditional methods, however. In the case of syntax, for example, corpus data may be crucial, despite the

combinatorial challenge. For example, Labov (1975, 1996) reported that it is possible for speakers to reject a construction in their overt judgments, while still systematically producing it themselves. Moreover, interpreting the acceptability of a syntactic construction depends on first finding appropriate tokens for testing. A corpus search sometimes reveals that a construction at first deemed impossible is actually attested in production, and often the discovered examples are not idiosyncratic, but generally acceptable to native speakers (e.g. Her 2009, who finds many plausible-sounding corpus examples that violate Chinese syntactic generalizations posited by Huang, Li, and Li 2009). Online, non-judgment-based experiments also have their place in theoretical syntax, not only as a means for studying the relationship between grammar and processing (e.g. Hawkins 2004, Phillips and Wagers 2007) but also as a check on the reliability and validity of judgments by probing inside the judgment-making process itself (e.g. Bornkessel-Schlesewsky and Schlesewsky 2007).

Relying solely on dictionary data in phonology is, if anything, even more risky, given the possibility for rote memorization. This realization has led to the mainstreaming in theoretical phonology (more so than in syntax) of "non-traditional" data sources. Alternative methods include the analysis of natural speech errors, which create novel forms that could not have been memorized (e.g. Fromkin 1971), experimentally collected judgments (e.g. Frisch and Zawaydeh 2001) and other offline tasks (Derwing and de Almeida 2009), and the phonetic analyses of speech articulation, which may reveal grammar-like (i.e. stable and partly learned) structure missed in informal transcriptions (Solé, Beddor, and Ohala 2007).

While making improvements in the collection and analysis of linguistic observations, however, we should not neglect the other end of the inferential chain, namely our models of mental grammar itself. Linguists already tend to have more detailed models of grammar than psychologists do for many of the mental entities and systems that they study, but grammatical claims are often difficult to test directly (e.g. the intermediate steps in derivational theories of phonology or syntax). Yet even the most abstract theory of grammar may have testable empirical consequences if they are described precisely enough. If the theory is multifaceted, as linguistic theories tend to be, generating quantitatively precise predictions may require computer modeling to keep track of the interactions of multiple components. Computational models have already become mainstream in the phonology literature, particularly constraint-based models of phonotactics, variation, and phonetic influences (e.g. Boersma & Hamann 2008, Coetzee & Pater 2008, Hayes & Wilson 2008). Computational models of syntax have so far been limited primarily to language acquisition (e.g. Waterfall et al. 2010), but the potential for progress here is great.

Quantification, experimentation, corpus analysis, and computer modeling are improvements on traditional grammatical methodologies not merely because of their origins in more "empirical" disciplines like phonetics and experimental psychology. Grammarians have different interests from researchers working on such topics as psychoacoustics and

parsing strategies, and so the types of data relevant to their hypotheses, and the methods used to collect and analyze them, need not be the same as in these other fields. Rather, the real advantage is that going beyond the traditional methods permits grammarians to have greater confidence in the strength of the inferential chain between observation and grammatical hypothesis. Just as in any science, strengthening this chain is a gradual, perhaps never-ending process, with each step revealing something new and important in the overall system of mind and behavior, wherein (somewhere, somehow) lurks the elusive mental grammar.

4. Preview

The future of grammatical research is not only bright and promising, but open-ended. Consensus over the nature of grammar will not emerge through a restriction of methodological tools, but through growth in our understanding of how data from a wide variety of sources fit together into a coherent whole, each linked to the multifaceted thing (whatever that turns out to be) at the core of it all. Hence I follow my sketch of methodological issues not with conclusions (tentative, dogmatic, or otherwise), but with a preview of coming attractions.

Following this overview chapter, Part 1 (Methodological issues) continues with three chapters introducing each of the key elements of an improved grammatical methodology: quantification, experimentation, corpus analysis, and computer modeling.

In chapter 2 ("Empirical methods in phonological research"), Michael Hammond guides the reader through a tutorial on empirical methods in theoretical phonology. Exemplifying the mainstreaming of both quantitative corpus analysis and experimental methods in theoretical phonology, Hammond demonstrates both types of techniques for testing phonological hypotheses, and illustrates them with a series of simple (and readily replicable) studies on English phonotactics. The free software tools used in the studies are explained as well, including R, the free statistics package widely used in quantitative linguistics (Baayen 2008, Johnson 2008).

In chapter 3 ("Doing experimental syntax: Bridging the gap between syntactic questions and well-designed questionnaires"), Wayne Cowart presents a tutorial on designing syntactic judgment experiments. Updating and extending a portion of Cowart (1997), the chapter provides a step-by-step guide for applying standard psycholinguistic methods to the primary data source in theoretical syntax, showing how to prepare the stimulus lists needed to test the generalizability of syntactic claims across both speakers and sentences, and how to distinguish effects due to stable grammar from those due to the fleeting influences that one sentence in a list may have on another. This lucid and thorough tutorial should make it possible for syntacticians with no prior psycholinguistic experience to get a judgment experiment ready in a matter of hours, without the need for any special software beyond what

they probably already have on their computers.

In chapter 4 ("Power-law distribution in morphological productivity: A statistical analysis of Chinese compounds"), Chao-Jan Chen applies a computational modeling approach to the corpus analysis of Chinese compounding. Unlike some corpus-based morphological studies which merely compare the relative rates of productivity of different morphemes (e.g. Baayen and Renouf 1996), Chen uses the corpus data to make an explicit comparison between a grammatical model in which compounds are built according to an abstract rule (COMPOUND → ROOT + ROOT) and an analogical model in which a root that already appears in many words combines with other roots more readily than roots appearing in only a few words. Based on the quantitative distribution of root productivity observed in a large Chinese corpus, Chen finds evidence in favor of the analogical model.

In Part 2 (Theoretical considerations), two chapters highlight the challenges one faces when attempting to forge links between hypothesized grammar and empirical observation.

In chapter 5 ("Distinguishing grammatical and processing explanations of syntactic acceptability"), Chien-Jer Charles Lin demonstrates how syntactic research can benefit from paying closer attention to extra-grammatical processing factors. Distinguishing the influence of such factors on sentence acceptability from the effects of grammar itself is not at all straightforward, even when syntacticians go beyond informal judgments to carefully designed experiments following standard psycholinguistic protocols. Lin reviews recent findings on sentence processing in English and Chinese that show that acceptability judgments can be systematically affected by a wide range of extra-grammatical factors. Indeed, there are cases where the line between grammar and processing remains difficult to draw. One striking example concerns the acceptability of resumptive pronouns in Chinese, which seems to be governed primarily by processability (with effects opposite to those observed in languages like English; Alexopoulou and Keller 2007), and yet there are also contexts in which grammar itself seems to play the decisive role.

In chapter 6 ("Testing phonological grammars with lexical data"), James Myers shows how the traditional word-list-based approach to theoretical phonology can be converted into quantitative corpus analysis while still working within Optimality Theory, currently the dominant framework in theoretical phonology (Prince and Smolensky 2004) and without requiring the researcher to be particularly experienced with statistical software (explicit instructions are given for building the necessary models in R). Myers explains the mathematical and philosophical basis of this approach, and then applies it in an analysis of the Formosan language Pazih. This analysis also addresses a common challenge in phonology, namely, how to interpret a quasi-systematic pattern in a language where speakers are not available for experimental testing.

Part 3 (Applications) presents two case studies illustrating the power of the new tools of grammatical science.

In chapter 7 ("What kind of thing is a coordinate?"), Cowart and Dana McDaniel adopt the experimental approach to syntactic judgments described in Cowart's tutorial in order to test a hypothesis straddling the traditional border between linguistics and cognitive psychology. Noting that coordinate structures are odd when compared with other syntactic constructions, they consider the possibility that this is because they are beyond "narrow syntax," with conjuncts forming an aggregate rather than a hierarchical structure. The authors point out that this hypothesis is already consistent with some observations made on the basis of traditional informal acceptability judgments, but certain crucial predictions are too subtle to test informally. In particular, the contrast between hierarchical versus non-hierarchical structure is expected to interact with judgments relating to subject-verb agreement, which are notoriously delicate and variable. Yet with the reduction in noise and bias made possible through carefully designed and analyzed experiments, the authors are able to show consistent judgment patterns supportive of their hypothesis.

In chapter 8 ("Corpus data vs. experiments in English phonotactics"), Hammond applies both corpus-based and experimental methods, as introduced in his tutorial, to a more subtle phonotactic issue (building on observations made in Hammond 1999). Here corpus-based and experimental data are not used merely to complement each other, but also to test hypotheses about the origins of phonotactic acceptability judgments. If judgments of nonwords are based solely on superficial analogy with existing lexical items, there should be a close correlation between such judgments and the patterns extracted (in a superficial, analogy-like way) from a database of attested words. By contrast, if judgments reflect an abstract grammar constrained by universal principles, they should show patterns that cannot be predicted by analogy alone. Hammond finds that nonword judgments by native speakers of English do indeed remain sensitive to universal constraints on syllable structure even when analogical similarity to existing words is factored out. Moreover, he shows how both types of effects may be modeled in terms of simple computational systems.

Hammond calls his tutorial an "invitation", as he aims to offer a range of tools and techniques that readers are free to use, reject, or extend for their own purposes. This is an excellent way to describe the anti-dogmatic call for methodological freedom espoused by all of the contributions in this book. I hope that our invitation to the ever-widening vistas of linguistic methodology will help inspire greater empirical robustness and theoretical coherence in the reader's own research.

References

Albright, Adam, and Bruce Hayes. 2003. Rules vs. analogy in English past tenses: Acomputational/experimental study. *Cognition* 90:119-161.

Alexopoulou, Theodora, and Frank Keller. 2007. Locality, cyclicity and resumption: at the

interface between the grammar and the human sentence processor. *Language* 83:110-60.

Aronoff, Mark, and Sungeun Cho. 2001. The semantics of *-ship* suffixation. *Linguistic Inquiry* 32.1:167-173.

Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge, UK: Cambridge University Press.

Baayen, R. Harald, and Antoinette Renouf. 1996. Chronicling the *Times*: Productive lexical innovations in an English newspaper. *Language* 72:69-96.

Bailey, Todd M., and Ulrike Hahn. 2001. Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory & Language* 44:569-591.

Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge, UK: Cambridge University Press.

Bloomfield, Leonard. 1914. *An Introduction to the Study of Language*. New York: Holt.

Bloomfield, Leonard. 1933. *Language*. New York: Holt, Rinehart & Winston.

Boersma, Paul, and Silke Hamann. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25:217-270.

Bornkessel-Schlesewsky, Ina, and Matthias Schlesewsky. 2007. The wolf in sheep's clothing: Against a new judgement-driven imperialism. *Theoretical Linguistics* 33:319-33.

Borsley, Robert D. 2005. Introduction. *Lingua* 115.11, ed. by Borsley, Robert D., 1475-1480. Special issue on data in theoretical linguistics.

Botha, Rudolf P. 1982. On the "Galilean style" of linguistic inquiry. *Lingua* 58:1-50.

Botha, Rudolf P. 1989. *Challenging Chomsky: The Generative Garden Game*. Oxford: Oxford University Press.

Bromberger, Sylvain, and Morris Halle. 1989. Why phonology is different. *Linguistic Inquiry* 20.1:51-70.

Cardona, George. 1994. Indian linguistics. *History of Linguistics: Volume I: The Eastern Traditions of Linguistics*, ed. by Giulio Lepschy, 25-60. London: Longman.

Carlson, Greg. 2003. On the notion 'showing something'. *The Nature of Explanation in Linguistic Theory*, ed. by John Moore and Maria Polinsky, 69-82. Stanford, California: CSLI Publications.

Carroll, John M., Thomas G. Bever, and Chava R. Pollack. 1981. The non-uniqueness of linguistic intuitions. *Language* 57:368-383.

Chalmers, A. F. 1999. *What Is This Thing Called Science?* 3rd ed. Indianapolis, IN: Hackett Publishing.

Chomsky, Noam. 1957. *Syntactic structures*. The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1980. *Rules and Representations*. New York: Columbia University Press.

Chomsky, Noam. 2002. *On Nature and Language*. Cambridge, UK: Cambridge University Press.

Chomsky, Noam, and Morris Halle. 1968. *The Sound Pattern of English*. New York: Harper and Row. Reprinted 1990, Cambridge, MA: MIT Press.

Churchland, Patricia S. 2002. *Brain-wise: Studies in Neurophilosophy*. Cambridge, MA: MIT Press.

Civil, Miguel. 1994. Sumerian. *History of Linguistics: Volume I: The Eastern Traditions of Linguistics*, ed. by Giulio Lepschy, 76-87. London: Longman.

Cleland, Carol E. 2002. Methodological and epistemic differences between historical science and experimental science. *Philosophy of Science* 69:474-496.

Coetzee, Andries W., and Joe Pater. 2008. Weighted constraints and gradient restrictions on place co-occurrence in Muna and Arabic. *Natural Language and Linguistic Theory* 26. 2: 289-337.

Coetzee, Andries W., René Kager, and Joe Pater. 2009. Introduction: Phonological models and experimental data. *Phonology* 26:1-8.

Cohen, I. Bernard. 1985. *Revolution in Science*. Cambridge, MA: Harvard University Press.

Cole, Jennifer, and José Ignacio Hualde. (Eds.) 2007. *Laboratory Phonology* 9. Berlin: Mouton de Gruyter.

Cowart, Wayne. 1997. *Experimental Syntax: Applying Objective Methods to Sentence Judgments*. London: Sage Publications.

Derwing, Bruce L, and Roberto G. de Almeida. 2009. Non-chronometric experiments in linguistics. *Experimental and Quantitative Linguistics*, ed. by David Eddington. Munich: Lincom.

Derwing, Bruce L., and William J. Baker. 1978. On the re-integration of linguistics and psychology. *Recent Advances in the Psychology of Language: Formal and Experimental Approaches*, ed. by R. N. Campbell and P. T. Smith, 193-218. New York: Plenum Press.

Devitt, Michael. 2006. *Ignorance of Language*. Oxford: Oxford University Press.

Di Sciullo, Anna Maria, and Edwin Williams. 1987. *On the Definition of Word*. Cambridge, MA: MIT Press.

Downer, G. B. 1963. Traditional Chinese phonology. *Transactions of the Philological Society* 62.1:127-142.

Edelman, Shimon, and Morten H. Christiansen. 2003. How seriously should we take Minimalist syntax? *Trends in Cognitive Science* 7.2:60-61.

Everaert, Martin, and Simon Musgrave. (Eds.) 2009. *The Use of Databases in Cross-linguistic Studies: Empirical Approaches to Language Typology*. Berlin: Mouton de Gruyter.

Everaert, Martin, Tom Lentz, Hannah De Mulder, Øystein Nilsen, and Arjen Zondervan (Eds.) 2010. *The Linguistics Enterprise: From Knowledge of Language to Knowledge in Linguistics*. Amsterdam: John Benjamins Publishing Company.

Fanselow, Gisbert, Caroline Fery, Matthias Schlesewsky, and Ralf Vogel. (Eds.) 2006.

Gradience in Grammar: Generative Perspectives. Oxford: Oxford University Press.

Featherston, Sam. 2007. Data in generative grammar: The stick and the carrot. *Theoretical Linguistics* 33.3:269-318.

Featherston, Sam, and Wolfgang Sternefeld. (Eds.) 2007. *Roots: Linguistics in Search of Its Evidential Base*. Berlin: Mouton de Gruyter.

Featherston, Sam, and Susanne Winkler. (Eds.) 2009. *The Fruits of Empirical Linguistics: Volume 1: Process*. Berlin: Mouton de Gruyter.

Frisch, Stefan A., and Bushra Adnan Zawaydeh. 2001. The psychological reality of OCP-Place in Arabic. *Language* 77.1:91-106.

Fromkin, Victoria A. 1971. The non-anomalous nature of anomalous utterances. *Language* 47:27-52.

Gillies, Donald. 1998. The Duhem Thesis and the Quine Thesis. *Philosophy of Science: The Central Issues*, ed. by Martin Curd and J. A. Cover, 302-319. New York: Norton.

Gonzalez-Marquez, Monica, Irene Mittleberg, Seana Coulson, and Michael J. Spivey. (Eds.) 2007. *Methods in Cognitive Linguistics*. Amsterdam: John Benjamins.

Gries, Stefan Th. 2009. *Quantitative Corpus Linguistics with R: A Practical Introduction*. New York: Routledge.

Gussenhoven, Carlos, and Haike Jacobs. 2011. *Understanding Phonology*, 3rd ed. London: Arnold.

Halle, Morris, and K. P. Mohanan. 1985. Segmental phonology of Modern English. *Linguistic Inquiry* 16:57-116.

Hammond, Michael. 1999. *The Phonology of English*. Oxford: Oxford University Press.

Hansson, Gunnar Ólafur. 2008. Diachronic explanations of sound patterns. *Language and Linguistics Compass* 2.5:859-893.

Hawkins, John A. 2004. *Efficiency and Complexity in Grammars*. Oxford: Oxford University Press.

Hayes, Bruce. 2000. Gradient well-formedness in Optimality Theory. *Optimality Theory: Phonology, Syntax, and Acquisition*, ed. by J. Dekkers, F. van der Leeuw, and J. van de Weijer, 88-120. Oxford: Oxford University Press.

Hayes, Bruce, and Colin Wilson. 2008. A maximum entropy model of phonotactics and phonotactic learning. *Linguistic Inquiry* 39.3:379-440.

Her, One-Soon. 2009. Unifying the long passive and the short passive: On the bei construction in Taiwan Mandarin. *Language and Linguistics* 10.3:421-470.

Huang, James. 1982. *Logical Relations in Chinese and the Theory of Grammar*. MIT dissertation.

Huang, James C.-T., Y.-H. Audrey Li, and Yafei Li. 2009. *The Syntax of Chinese*. Cambridge, UK: Cambridge University Press.

Johnson, Keith. 2008. *Quantitative Methods in Linguistics*. Oxford: Blackwell Publishers.

Kennedy, Graeme. 1998. *An Introduction to Corpus Linguistics*. New York: Addison Wesley Longman.

Kenstowicz, Michael, and Charles Kisseberth. 1979. *Generative Phonology: Description and Theory.* New York: Academic Press.

Kepser, Stephan, and Marga Reis. (Eds.) 2005. *Linguistic Evidence: Empirical, Theoretical and Computational Perspectives*. Berlin: Mouton de Gruyter.

Kingston, John, and Mary E. Beckman. (Eds.) 1990. *Papers in Laboratory Phonology 1: Between the Grammar and Physics of Speech*. Cambridge, UK: Cambridge University Press.

Krifka, Manfred. (Eds.) 2007. *Theoretical Linguistics* 33.3. Special issue on data in generative grammar.

Kuhn, Thomas S. 1970. *The Structure of Scientific Revolution*, 2nd ed. enlarged. Chicago: University of Chicago Press.

Labov, William. 1975. Empirical foundations of linguistic theory. *The Scope of American Linguistics*, ed. by R. Austerlitz, 77-113. Lisse: The Peter de Ridder Press.

Labov, William. 1996. When intuitions fail. *Chicago Linguistics Society* 32:77-105.

Leben, William R. 1978. The representation of tone. *Tone: A Linguistic Survey*, eds. by V. A. Fromkin, 177-219. New York: Academic Press.

Luka, Barbara J., and Lawrence W. Barsalou. 2005. Structural facilitation: Mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language* 52:436-59.

Manning, Christopher D., and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. Cambridge, MA: MIT Press.

Matthews, Robert J. 1991. Psychological reality of grammars. *The Chomskyan Turn*, ed. by Asa Kasher, 182-199. Oxford: Basil Blackwell.

Meibauer, Jörg, and Markus Steinbach. (Eds.) 2011. *Experimental Pragmatics/Semantics*. Amsterdam: John Benjamins Publishing Company.

Millikan, Ruth Garrett. 1995. *White Queen Psychology and Other Essays for Alice*. Cambridge, MA: MIT Press.

Myers, James. 2007. Generative morphology as psycholinguistics. *The Mental Lexicon: Core Perspectives*, ed. by G. Jarema and G. Libben, 105-128. Amsterdam: Elsevier.

Myers, James. 2009a. Syntactic judgment experiments. *Language and Linguistics Compass* 3. 1:406-423.

Myers, James. 2009b. The design and analysis of small-scale syntactic judgment experiments. *Lingua* 119:425-444.

Myers, James. 2009c. Automated collection and analysis of phonological data. *The Fruits of Empirical Linguistics: Volume 1: Process*, ed. by S. Featherston and S. Winkler, 151-176. Berlin: Mouton de Gruyter.

Nagata, Hiroshi. 1989. Repetition effect in judgments of grammaticality of sentences: Examination with ungrammatical sentences. *Perceptual and Motor Skills* 68:275-282.

Nagata, Hiroshi. 2003. Judgments of grammaticality of Japanese sentences violating the principle of Full Interpretation. *Journal of Psycholinguistic Research* 32:693-709.

Neeleman, Ad, and Hans van de Koot. 2010. Theoretical validity and psychological reality of the grammatical code. *The Linguistics Enterprise: From Knowledge of Language to Knowledge in Linguistics*, ed. by Martin Everaert, Tom Lentz, Hannah De Mulder, and Øysterin Nilsen, 183-212. Amsterdam: John Benjamins.

Newmeyer, Frederick J. 1983. *Grammatical Theory: Its Limits and Its Possibilities*. Chicago: University of Chicago Press.

Newmeyer, Frederick J. 2007. Commentary on Sam Featherston, 'Data in generative grammar: The stick and the carrot'. *Theoretical Linguistics* 33:395-99.

Ohala, John J. 1986. Consumer's guide to evidence in phonology. *Phonology Yearbook* 3: 3-26.

Ohala, John J. and Jeri J. Jaeger. (Eds.) 1986. *Experimental Phonology*. New York: Academic Press.

Penke, Martina, and Anette Rosenbach. (Eds.) 2007a. *What Counts as Evidence in Linguistics: The Case of Innateness*. Amsterdam: John Benjamins.

Penke, Martina, and Anette Rosenbach. 2007b. What counts as evidence in linguistics? An introduction. *What Counts as Evidence in Linguistics: The Case of Innateness*, ed. by Martina Penke and Anette Rosenbach, 1-49. Amsterdam: John Benjamins.

Phillips, Colin. 2009. Should we impeach armchair linguists? *Japanese/Korean Linguistics* 17, ed. by Shoichi Iwasaki. Palo Alto, CA: Center for the Study of Language and Information.

Phillips, Colin, and Howard Lasnik. 2003. Linguistics and empirical evidence: Reply to Edelman and Christiansen. *Trends in Cognitive Science* 7:61-62.

Phillips, Colin, and Matthew Wagers. 2007. Relating structure and time in linguistics and psycholinguistics. *Oxford Handbook of Psycholinguistics*, ed. by G. Gaskell, 739-756. Oxford: Oxford University Press.

Prince, Alan, and Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Oxford: Blackwell.

Pullum, Geoffrey K. 1983. The revenge of the methodological moaners. *Natural Language and Linguistic Theory* 4, 583-588. Reprinted 1991 in Geoffrey K. Pullum, *The Great Eskimo Vocabulary Hoax and Other Irreverent Essays on the Study of Language*, 123-130. Chicago: University of Chicago Press.

Rice, Sally, and John Newman. (Eds.) 2010. *Empirical and Experimental Methods in Cognitive/Functional Research*. Stanford, California: Center for the Study of Language and Information.

Ritter, Nancy A. (Ed.) 2002. *The Linguistic Review* 19.1-2. Special issue on the poverty of the stimulus argument.

Sampson, Geoffrey R. 2001. *Empirical Linguistics*. London: Continuum.

Sampson, Geoffrey R. 2007. Grammar without grammaticality. *Corpus Linguistics and Linguistic Theory* 3:1-32.

Sapir, Edward. 1949. The psychological reality of phonemes. *Selected Writings of Edward Sapir in Language, Culture, and Personality*, D. Mandelbaum, 46-60. Berkeley: University of California Press. Originally published in French as Sapir, E. 1933. La réalité psychologique de phonèmes. *Journal de Psychologie Normale et Pathologique* 30:247-265.

Schlinger, Hank D. 2004. Why psychology hasn't kept its promises. *Journal of Mind and Behavior* 25:123-144.

Schütze, Carson T. 1996. *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology*. Chicago: University of Chicago Press.

Schütze, Carson T. 2011. Linguistic evidence and grammatical theory. *Wiley Interdisciplinary Reviews: Cognitive Science* 2.2:206-221.

Seuren, Pieter A. M. 1998. *Western Linguistics: An Historical Introduction*. Oxford: Blackwell Publishers.

Shapin, Steven. 1996. *The Scientific Revolution*. Chicago: University of Chicago Press.

Solé, Maria-Josep, Patrice Speeter Beddor, and Manjari Ohala. (Eds.) 2007. *Experimental Approaches to Phonology*. Oxford: Oxford University Press.

Spencer, Andrew. 1991. *Morphological Theory: An Introduction to Word Structure in Generative Grammar*. Oxford: Basil Blackwell.

Sproat, Richard and Osamu Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics* 21:291-311.

Sprouse, Jon. 2007. Continuous acceptability, categorical grammaticality, and experimental syntax. *Biolinguistics* 1:117-28.

Sprouse, Jon. 2008. The differential sensitivity of acceptability to processing effects. *Linguistic Inquiry* 39:686-94.

Sprouse, Jon. 2009. Revisiting satiation: Evidence for an equalization response strategy. *Linguistic Inquiry* 40.2:329-341.

Sprouse, Jon. 2011. A test of the cognitive assumptions of magnitude estimation: Commutativity does not hold for acceptability judgments. *Language* 87.2:274-288.

Stefanowitsch, Anatol. (Ed.) 2007. *Corpus Linguistics and Linguistic Theory* 3.1. Special issue on target article Sampson (2007).

Stefanowitsch, Anatol, and Stefan Th. Gries. (Eds.) 2007. *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*. Berlin: Mouton de Gruyter.

Uffmann, Christian. 2006. Epenthetic vowel quality in loanwords: Empirical and formal

issues. *Lingua* 116:1079-1111.

Uttal, William R. 2007. *The Immeasurable Mind: The Real Science of Psychology*. Amherst, NY: Prometheus Books.

Waterfall, Heidi R., Ben Sandbank, Luca Onnis, and Shimon Edelman. 2010. An empirical generative framework for computational modeling of language acquisition. *Journal of Child Language* 37:671-703.

Winkler, Susanne, and Sam Featherston. (Eds.) 2009. *The Fruits of Empirical Linguistics: Volume 2: Progress*. Berlin: Mouton de Gruyter.

Xu, Liejiong. 1990. Remarks on LF Movement in Chinese questions. *Linguistics* 28:355-382.

Yule, John-David. (Ed.) 1978. *Concise Encyclopedia of the Sciences*. New York: Van Nostrand Reinhold Company.

Zoll, Cheryl. 2003. Optimal tone mapping. *Linguistic Inquiry* 34.2:225-268.

Zuraw, Kie. 2007. The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog infixation. *Language* 83.2:277-316.